REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	12/23/96	Final - 5/1/94 - 4/30/96

4. TITLE AND SUBTITLE

SPACE PERCEPTION WITH NORMAL AND PROSTHETIC VISION
(Space Perception)

5. FUNDING NUMBERS

G - F49620-94-1-0262

61102F

2313/CS

6. AUTHOR(S)

Itzhak Hadani
Bela Julesz

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Rutgers, The State University of New Jersey
P.O. Box 1089
Piscataway, NJ 08854-1089

8. PERFORMING ORGANIZATION REPORT NUMBER

4-26395

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Captain William Roach
AFOSR - NL 1100
110 Duncan Avenue - Suite B115
Bolling AFB, DC 20332-8080

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

93-NL-165

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This report encompasses: a) A unique and metric solution for the differential equations that specify the optic-flow of monocularly navigating observer. b) A report on a correlation between individual differences in interocular distance and registered depth in random dot stereogram with and without pedestal disparities. c) Expansion on a navigational approach to space perception - SPIN theory - which suggests that object constancy is obtained during fixations by pure passive navigation computations, and across saccades by a combination of ocular and vestibular signals. It is suggested that the VOR constraints the rotational and velocity components of the eye to be perpendicular. d) Analysis of the degree of uncertainty offered by the inferential, direct, and computational approaches in cognitive psychology as illustrated by their window metaphors. Visual stability in normal and prosthetic vision is examined and leads to newly stated magnification and distance paradoxes. A telescope metaphor, which is a modified Mach-Gibson visual-ego metaphor with a zooming feature, is suggested as a model that can resolve the paradoxes. e) A computer system which simultaneously displays motion parallax yoked to head movement and binocular disparity, with measurements of the virtual parallax evoked by head movements in static RDS.

14. SUBJECT TERMS

Computation Vision, Scale Ambiguity, Depth Constancy,
SPIN Theory, Inter-Ocular Distance, Prosthetic Vision

15. NUMBER OF PAGES

194

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
U	U	U	U

SPACE PERCEPTION WITH NORMAL AND PROSTHETIC VISION

Scientific Report

Submitted By

Itzhak Hadani and Bela Julesz

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. Government.

December 1996

SPACE PERCEPTION WITH NORMAL AND PROSTHETIC VISION

ABSTRACT

This report encompasses: a) A unique and metric solution for the differential equations that specify the optic-flow of monocularly navigating observer. b) A report on a correlation between individual differences in interocular distance and registered depth in random dot stereogram with and without pedestal disparities. c) Expansion on a navigational approach to space perception - SPIN theory - which suggests that object constancy is obtained during fixations by pure passive navigation computations, and across saccades by a combination of ocular and vestibular signals. It is suggested that the VOR constraints the rotational and velocity components of the eye to be perpendicular. d) Analysis of the degree of uncertainty offered by the inferential, direct, and computational approaches in cognitive psychology as illustrated by their window metaphors. Visual stability in normal and prosthetic vision is examined and leads to newly stated magnification and distance paradoxes. A telescope metaphor, which is a modified Mach-Gibson visual-ego metaphor with a zooming feature, is suggested as a model that can resolve the paradoxes. e) A computer system which simultaneously displays motion parallax yoked to head movement and binocular disparity, with measurements of the virtual parallax evoked by head movements in static RDS.

Table of Contents

		Page
Chapter 1	Introduction	1
Chapter 2	Passive navigation for an eye in 6 degrees of freedom	9
Chapter 3	The effect of inter-ocular distance on registered depth in RDS with different pedestal disparities	41
Appendix A	The SPIN theory - A navigational approach to space perception	
Appendix B	Perceptual constancy and the mind's eye looking through a telescope	
Appendix C	Simple integrative method for presenting head contingent motion parallax and disparity cues on Intel processor based PCs	
Appendix D	Master's thesis: SPIN theory and indeterminate scale problem	

CHAPTER 1: INTRODUCTION

Fundamental problems in space perception are: a) How can a mobile observer navigate in space on the basis of retinal and extraretinal signals? and b) How are the perceptual constancies of position size and shape obtained during fixation and across eye and head movements? These questions are accounted for by the SPIN theory - a navigational approach to space perception - that served as a theoretical framework to the present work. The goals of this work were: a) to solve the differential equations specifying the optic flow of an eye navigating in 6 degrees of freedom, and b) to expand theoretically on the SPIN theory including formalization of constraint(s) that are imposed by the laws of eye movements and the VOR, and the question of how passive system can determine perceptually whether its computations convey veridical depth.

The main thrust of the research was focused on the problem of passive navigation, e.g. how a system can navigate without any external information. To address this question computational models of vision are utilizing information conveyed by the optic flow and try to recover the egodistance and the 6 pure motion parameters (3 rotational components and 3 velocity components) of the eye in space. This is a two century old problem and so far led to indeterminate solutions which means that the problem can be solved only up to a scalar in the velocity vector. This restriction is named nowadays as the scale ambiguity problem. In contrast, the present work discovered a unique metric solution to the problem. It is presented in chapter 2 and involves a simultaneous solution of a system of 5 linear equations. Two main features made this solution possible. The first is the application of navigational considerations that added a constant integration factor that in our case turned out to be an arbitrary magnitude. the second is the positioning of eye system origin at the geometric center of the eye instead of at the pinhole. This feature added a inhomogeneous element to the coordinate's world points. We note that since the problem of depth from motion is isomorphous to depth from stereopsis and since the present solution utilizes triangulation, it may be applied to stereoscopic vision as well.

Chapter 2 represents only a small fraction of the computational work invested on this issue. A more comprehensive account is given in Appendix D of this report which comprises the M.A. thesis of Alex Kononov (who did his work within the framework of this project). In fact, three different approaches were applied to solve the differential equations of the optic flow. These were the differential non-linear, the differential linear and the discrete approaches. A major achievement of the present work is the ability to solve, in all three approaches, for the rotational component of the eye irrespective of the distance of points and velocity. The solution in the differential linear is given in chapter 2 and the solution in the discrete case for points along the same meridian is given in Appendix D of the appended Kononov's thesis.

Chapter 3 presents a study that focuses on one narrow aspect of stereoscopic depth constancy i.e., the question of how the metric measure of interocular distance does affect depth appreciation in random dot stereogram. Geometry of binocular vision shows that disparity is scaled by the inverse of the square of the viewing distance. This is what is known as the inverse square law which also sets the relation between the latter magnitudes and depth and interocular distance (IOD). While there are many studies on the effects of viewing distance and disparity on perceived depth, little work has been done on the effect of IOD. The latter effect can be investigated either by optical modification of the natural IOD, or by using individual differences in IOD as is done in the present study. The evaluation of the effect of this metric measure has theoretical and practical implications. In four correlational experiments, a large pool of observers ranging in age from 5 to 63, males and females, with or without corrected vision, with IOD ranging from 4.9 cm to 7.3 cm, were presented with the same stereogram depicting a cyclopean square hovering over the background. The stereogram was viewed by all observers from the same viewing distance. In different experiments the background had -2, 0, +2 pixel units disparities while the square had always a positive disparity. Observers had to align a real probe, viewed binocularly with the perceived depth of the square and the background. The two alignments (measured in cm), which we refer to as registered depth, were plotted against the observers' IOD as a quasi-independent

variable. The individual depth judgments fell nicely on the theoretical curves predicted by the inverse square law and without the need to fit a free parameter. The theoretical implications of these findings to the issue of perceptual constancy as well their practical implications to the design of virtual reality, head mounted, and other stereoscopic displays are discussed.

Appendix A presents the SPIN theory-a navigational approach to space perception - that provides a computational answer to the perceptual constancy problem. The theory postulates that mental representation is exocentric and that in normal and uncorrected vision objects are perceived, within measurement error, at their veridical location. There is a distinction between passive navigation where retinal signals are the sole source of information, and active navigation where extraretinal signals are combined with retinal signals. The theory postulates that the optic flow has all the necessary information for 3-D and self motion reconstruction. During saccades extraretinal signals are utilized to obtain the continuity of visual experience. The theory takes into account the Listing and Donders' Laws of eye movements and the Vestibulo-ocular reflex and suggests that the latter can be characterized by constraining the rotational and velocity vectors to be perpendicular. In that case, the eye motion in space becomes purely rotational and the computational procedures may be simplified. For scaling the reconstructed world it is suggested that the system utilizes three intrinsic scales. For monocular vision the scale is the radius of the eyeball. For binocular vision the scale is the Inter-ocular distance. For active navigation the scale is the distance between the eye system origin and the head system origin. Some empirical support for the conjecture of the role of Interocular distance as an intrinsic scale in binocular vision is presented in chapter 3. The question of the impairments to veridical space perception caused by prosthetic vision is discussed in more details in Hadani (1991) in relation to Night Vision Goggles. In Appendix B, this question is discussed by referring to the effects of simple lens, and in chapter 3 it is discussed for the hyperstereo case created by the use of telestereoscope. The simulation of the navigational procedures of project-onto-the-retina and project-out-in-space are presented in Hadani et al. (1995). In principle, it is possible to use the computational model of chapter 2 to predict the

location in space that objects will be perceived when the observer utilizes prosthetic vision. These predictions have to consider one additional factor; the esoteric adaptation capacity. In our view the latter factor is a scalar magnitude and can be determined empirically.

Appendix B deals with the problem of veridicality of space perception in normal and prosthetic vision. Perceptual constancy of position size and shape during head and eye-movements, is a problem dealt with by the inferential and direct perception approaches, and more recently, by computational vision. Yet it is not fully understood. To explain the essence of visual perception, different approaches utilize window metaphors through which the hypothetical mind's eye observes the world. In the inferential approach the window, called Alberti's window, is located in space between the observer and the objects. In the Gibsonian direct perception approach the window, called the Visual Ego, is fixed to the head. In computational vision the "window" is fixed to the eye, but the nature of the effective window after the computations are carried out remains ambiguous and depends on whether one assumes relative or absolute depth perception. In Appendix B we examine the theoretical meaning of the various windows, and suggest two versions of the same window metaphor: One represents the perception of relative depth, and the second, absolute depth. For both versions the window can be regarded as an adaptive zooming telescope that preserves stability of a 3-D world when the effective magnification is 1:1 and when the images projected on the mind's eye are collimated. Thus, the mind's eye can be conceived as a fictitious observer (say, an array of spatiotopically mapped receptive fields) viewing the world via non-magnifying telescope, and the objects are seen at optical infinity (first view) or, at their veridical egodistance (second view). It is argued that the stability problem cannot be accounted for by the Alberti's window and when analyzed with the Visual Ego leads to perceptual puzzles which we name the "Magnification and Distance Paradoxes." These puzzles become more enigmatic when observers use head mounted optical devices, which modify the effective retinal magnification, visual stability, and egodistance. With simple observations that can be performed by any interested reader, we demonstrate that the perceptual paradoxes raised by the

use of head mounted devices may be resolved with either version of the telescope metaphor. In this appendix the differences between the three approaches to visual perception is analyzed in terms of the degree on uncertainty in mapping between distal stimulus and its mental representation. These differences are illustrated with the pertinent window metaphors.

Appendix C deals with the development of a computer system which simultaneously displays motion parallax yoked to head movement and binocular disparity, with some related psychophysical measurements. It is known that motion parallax that is yoked with head movements produces monocular depth perception in a random dot pattern. The existing displays required a sophisticated hardware system that requires costly hardware and electronic workshop facilities. In this chapter we show how such a display system can be produced with any Intel processor based PC with no additional cost for hardware. An adapted joystick sampled by standard game-port can serve as a provisional head movement sensor. Color monitor resolution in displaying motion was effectively enhanced 16 times by the use of anti-aliasing enabling to display array of 1000 random dots in real time (double buffer mode) with a refresh rate of 60 Hz and above. The color monitor enabled the use of anaglyph method, thus combining stereoscopic and monocular parallax manipulations on a single CRT display without refresh rate cost. The power of our display is demonstrated by psychophysical measurements in which three subjects annulled head movement contingent illusory parallax, evoked by static random dot stereogram, with real counter-parallax. The amount of the real counter-parallax required to annul the illusory stereoscopic parallax monotonically increased with disparity.

Chapters 2 and 3 are in progress. Appendix A was published in the 1995 special issue Spatial Orientation (AFOSR) of the *Journal of Vestibular Investigation* vol. 5, 6, 443-454. Appendix B is in press and will appear in *Constraining Cognitive Theories - Issues and Options*, Z. Pylyshyn (Ed.), Norwood: Albex Publishing. Appendix C is in press and will be published in a special issue on display systems in visual psychophysics of the *Journal of Spatial Vision*. In addition, a peer commentary entitled "Computational aspects of motion perception during self

motion" was published within the framework of this project in *Behavioral and Brain Sciences* 1994, 17, 319-320.

PERSONNEL SUPPORTED/ASSOCIATED

Dr. Itzhak Hadani	Principal Investigator
Dr. Bela Julesz	Senior Investigator
Dr. Harry L. Frisch	Consultant
Alex Kononov	M.A student (accomplished his thesis within this project. Title: SPIN theory and indeterminate scale problem)
Janos Szatmary	(Programmer, paid hourly)
Aravind K. Suri	(Programmer, paid hourly)
Boris Burakov	(Programmer, paid hourly)
Carol A. Esso	(Secretary, paid hourly)

INTERACTIONS

1. Hadani, I., The SPIN theory - A navigational approach to space perception, seminar, Weizmann Institute, Israel, 1/5/94.
2. Hadani, I., Passive navigation in 6 degrees of freedom, seminar, Technion-Israel Institute of Technology, 1/6/94.
3. Hadani, I., The SPIN theory - A navigational approach to space perception, seminar, Haifa University, 1/6/94.
4. Hadani, I., The SPIN theory - A navigational approach to space perception, AFOSR Conference on Spatial Orientation, San Antonio, TX, May 18-20, 1994.
5. Hadani, I., Space perception in navigation (SPIN) theory, Neuroinformatics Department, University of Bochum, Germany, August 31, 1994.

6. Hadani, I., A navigational approach to visual space perception, Institut für Arbeitsphysiologie, University of Dortmund, Germany, September 1, 1994.

7. Hadani, I., The SPIN theory, 17th European Conference on Visual Perception (ECVP), Eindhoven, The Netherlands, Sept. 4-8, 1994, *Perception*, 23 (supplement), 108-109, 1994.

8. Hadani, I. and Julesz, B., The effect of inter-ocular distance on perceived depth in RDS at different pedestal disparity. Investigative Ophthalmology and Visual Science, 36 (4), p.1740, Proc. of ARVO Meeting, May 14-19 1995.

NEW DISCOVERIES

Even though we think that some of our ideas are patentable, so far no steps were taken in this direction. One potential application of the formal model is to convert computationally the effective vantage point of a camera. If the images of a moving imaging device, for example that of a thermal imaging system of an helicopter, are sampled and the motion field is extracted, then the formal model can be applied to calculate the true location in space of objects in the field of view. If, in addition, the pilot's vantage point is known, then a new image of the reconstructed objects can be created which represents their view from the pilot's vantage point.. The latter images can then be presented to the pilot via his helmet display system and will provide him veridical perception of the objects. We also conceive the potential use of the formal model to the construction of optical range finder for a given point in the visual field, and/or whole image range finder.

REFERENCES

Hadani, I. (1991). Corneal lens goggle and visual space perception. Applied Optics, 30, 28, 4136-4147.

Hadani, I. (1995). The SPIN theory - A navigational approach to space perception. Journal of Vestibular Research, 5, 6, 443-454.

CHAPTER 2:
PASSIVE NAVIGATION FOR AN EYE IN 6 DEGREES OF FREEDOM

Itzhak Hadani, Alex Kononov and Harry L. Frisch

2.1. INTRODUCTION

One primary task of the visual system is to enable the organism to navigate in 3-D space, and everyday human experience and performance indicates that vision acquires information about the objects in the world with impressive speed, reliability, and generality. When the visual system relies only on retinal information, the spatial orientation capacity is called autonomous or passive navigation (Bruss & Horn, 1983; Horn, 1986). Navigation implies that the eye is constantly moving in space and changes its position relative to objects. Pure geometrical considerations predict that the retinal image of the objects is subject to considerable instability and to continual severe deformations of size and shape. However, these effects are not reflected in the introspective impression of the objects which attain constant position size and shape. This state of affairs raises the question: How can the capacity of passive navigation be carried out while preserving the perceptual constancies of objects? The common view, based on many theoretical analyses, regards this problem as indeterminate (see e.g. Tsai & Huang, 1985; Ullman, 1979; Bruss & Horn, 1983; Koenderink & van Doorn, 1991). In this respect the Space Perception In Navigation (SPIN) theory differs in that it considers project-onto-the-retina and project-out-in-space aspects of mental representation (Hadani, 1995). These considerations lead to a simple formal model of coordinate transformation. Earlier work (Hadani *et al.*, 1994) demonstrated the feasibility of a metric solution in the coordinate transformation model for an eye in pure rotation. Here we show a solution for the more general case of an eye in 6 degrees of freedom.

To see how the idea of coordinate transforation is applied to vision, imagine a world which is comprised of a single static object and a freely navigating eye. Consider two cartesian coordinate systems: one is attached to the object and the other is attached to the eye. Indeed, each object-point has a 3-D representation in both coordinate systems, one is time invariant and the other is time

variant. If the relative position between the two systems is known, the time varying representation of the object in the eye system can be translated, by linear transformation, into a representation in the world system. This enables an observer attached to the moving system to calculate and adopt the representation of the object in the static system. There is, though, one essential difference between visual reality and this ideal model. The eye has only a two-dimensional perspective representation of the object and the depth coordinate is missing. Thus, in order to obtain a world-centered representation, the distance (depth) of the objects have to be recovered too. The main question is whether an observer who is attached to the moving system can reconstruct from retinal signals all these magnitudes. This problem has a long history in optics and vision research (see e.g. Cutting, 1986; Koenderink & van Doorn, 1991). While recovering structure and motion parameters can be metrically solved for 4 non-coplanar points and three orthographic views (Ullman, 1979; Koenderink & van Doorn, 1991), it has no unique solution in perspective views. It is argued that since the moving observer measures only angles, there is no way to get scaled distances of objects or egomotion. In contrast, we show here that the coordinate transformation model can be realized by a passive visual system when navigational considerations are added.

Formally, the problem of passive navigation is posed as follows (Duric *et al.*, 1995): Given a sequence of images taken by a monocular observer undergoing continuous rigid motion in a static environment, recover the distance of objects and the translation and rotation parameters of the observer (egomotion). The most general case of the observer's motion is unrestricted rigid motion, which can be represented by six independent parameters: three rotations and three translations. Broadly speaking, 3-D reconstruction is a two step process. First, the system needs to compute the instantaneous image velocity field or the retinal optic flow, or to solve the correspondence problem (Ullman, 1979) e.g. to establish object identity between elements in successive images. Second, the system has to recover egomotion and distances of objects from the data. This paper is concerned with the second stage of the process assuming the image velocity field is given.

Passive navigation has been studied extensively in the past two decades (Hadani *et al.* 1978, 1980; Ullman, 1979; Longuet-Higgins & Prazdny, 1980; Tsai & Huang, 1985), and is still one of the major topics of interest in the field of machine vision (Fermuller & Aloimonos, 1995; Duric *et al.*, 1995) because in this field the inputs are perspective projections taken by a camera. Three types of approaches, the discrete, the continuous, and the least-squares, have been pursued in most of the earlier works. In the continuous (differential) approach, the optic flow is used to extract the navigational parameters and structure from a point and its nearest neighborhood (Longuet-Higgins & Prazdny, 1980). In the discrete approach, information at only a few points is used to determine structure and the motion parameters (Longuet-Higgins, 1981; Meiri, 1980; Nagel, 1981; Tsai & Huang, 1985). In the least-squares approach, motion parameters are found that are most consistent with the estimated image velocity over the entire image (Prazdny, 1981; Bruss & Horn, 1983). One of the most thorough studies on uniqueness of the solution was carried out by Tsai and Huang (1985). Utilizing the discrete approach, they have shown that the motion parameters can be determined up to a scaling factor in the translation vector. The latter shortcoming is common to all alternative navigational approaches and the problem is called the *indeterminate scale problem* (see e.g. Fermuller & Aloimonos, 1995 for recent restatement of the problem).

The indeterminate scale problem has created theoretical difficulty in understanding human space perception mainly because motion parallax is considered as a primary cue for metric depth. The problem was illustrated by Bruss and Horn (1983, p. 6) as follows: "Consider a surface S_2 which is a dilation by factor k of a surface S_1. Further, let two motions denoted by V_1 and V_2 have the same rotational component and let their translational components be proportional to each other by the same factor k (we shall say that V_1 and V_2 are *similar*). Then the optical flow generated by S_1 and V_1 is the same as the optical flow generated by S_2 and V_2." Thus, if a biological system

applies computations on the optic flow to recover its own motion parameters and the distance and structure of objects, then there is no way in which such a system could decide whether the object is "small and nearby" or "big but far away" (Ullman, 1979, p.199). Let us examine the theoretical implications of the latter statement. In the first place it means that for a given velocity field an infinite family of visual objects varying in size and correspondingly proportional infinite magnitudes of observer's velocities can be reconstructed from the same retinal flow. As a result, metric scale for egodistance perception and egomotion cannot be established. The current view is that veridical precepts of objects can be obtained only with additional external information; more specifically, by information on egomotion that is given by the vestibular system (see e.g. Landy *et al.*, 1995). Indeed, it is known that the perceptual mechanism has access to extraretinal signals - ocular and vestibular - about the motion of the eye in space (see e.g. Henn *et al.*, 1984) and can combine them with retinal information. However, it is argued that the known properties of extraretinal signals make them insufficient to accurately supplement retinal signals because of the physiological noise in the oculomotor system (MacKay, 1973) and the insensitivity of vestibular system to linear motion (Howard, 1986). Therefore, they cannot fully account for the high degree of visual stability that we normally experience. Thus, one need to turn to retinal signals as a reliable and accurate source of information for navigational capacities.

Without violating generality, and for the sake of abbreviation and mathematical simplicity, we chose to present the details of our derivations in the differential approach. However, it can be shown that the same results can be obtained in the discrete approach directly from the flow field (Kononov, 1996, Appendix D). But the latter approach involves long and cumbersome derivations. The differential analysis, on the other hand, has two alternative ramifications. One alternative assumes that objects in space are comprised of a continuous rigid surface and each point projects onto the retina a Dirac delta function. This case requires one to consider, in addition to distance, the unknown structure gradient of the surface. In practice, the gradient components can be eliminated in the partial spatial derivatives (see Appendix A in Hadani *et al.*, 1994) but also

leads to long and cumbersome derivations. The second alternative presented here considers projection of a single general point that has no structure and leads to much simpler expressions. It is assumed that every point in space projects onto the retina a point spread function (Hecht & Mintz, 1936). The point spread function creates for the point a small flow field in its neighborhood which enables estimation of the spatial derivatives. This approach also has the benefit of relaxing the rigidity assumption. The computational and psychophysical arguments justifying the need to account for the perception of a single point (say, a star) were presented at length elsewhere (Hadani *et al.*, 1994). The latter are based mainly on the substantial visual stability shown in autokinesis, e.g. when the input to the system is a single static point observed in complete darkness.

2.2. KINEMATICS OF RETINAL PROJECTION IN 6 DEGREES OF FREEDOM

A. Definition of notations

P, **Q**, **R**	position vectors in world system
p, **q**, **r**	position vectors in eye (moving) system
X, Y, Z	cartesian coordinates of a point in the world system
x, y, z	cartesian coordinates of a point in the eye system
θ, φ , ρ	retinal polar coordinates
θ_t, $\theta_t\theta$	partial derivatives with respect to the variables as denoted by the subscript
θ_0, θ_1	coordinates of point in temporal order (view) as denoted by the subscript
V	velocity vector in world system
v	velocity vector in eye system
$\boldsymbol{\omega}$	rotation vector in eye system
A, B	two components of $\boldsymbol{\omega}$ in new basis

D, E	two components of **v** in new basis
V_a, V_b	compound variables of A, B, C, D
Λ	a flow field position-invariant constant
$\underline{\mathbf{M}}$	matrix that transform vectors from world to eye coordinates
$\underline{\mathbf{M}}^T$	matrix that transform vectors from eye to world coordinates
$\underline{\mathbf{I}}$	identity matrix
κ,λ,μ	Euler's angles that specify the transformation matrix $\underline{M}$
c	radius of the eyeball
c	radius vector of the eyeball

B. Basic assumptions

1) The eyeball is considered as a rigid sphere with a radius c.

2) The eye optics is reduced to a pinhole, thus preserving the central projection (perspective) property of the retinal flow.

3) Each point in space projects on the retina a point spread function.

4) The retinal flow is twice differentiable with respect to space.

C. Model

Let X, Y, Z and x, y, z be two cartesian coordinate systems fixed with respect to space and the eyeball respectively as shown in fig. 1a. The eye system origin coincides with center of the sphere. The optical axis is defined as a line passing through the center of the sphere and the pinhole coinciding with the y axis.

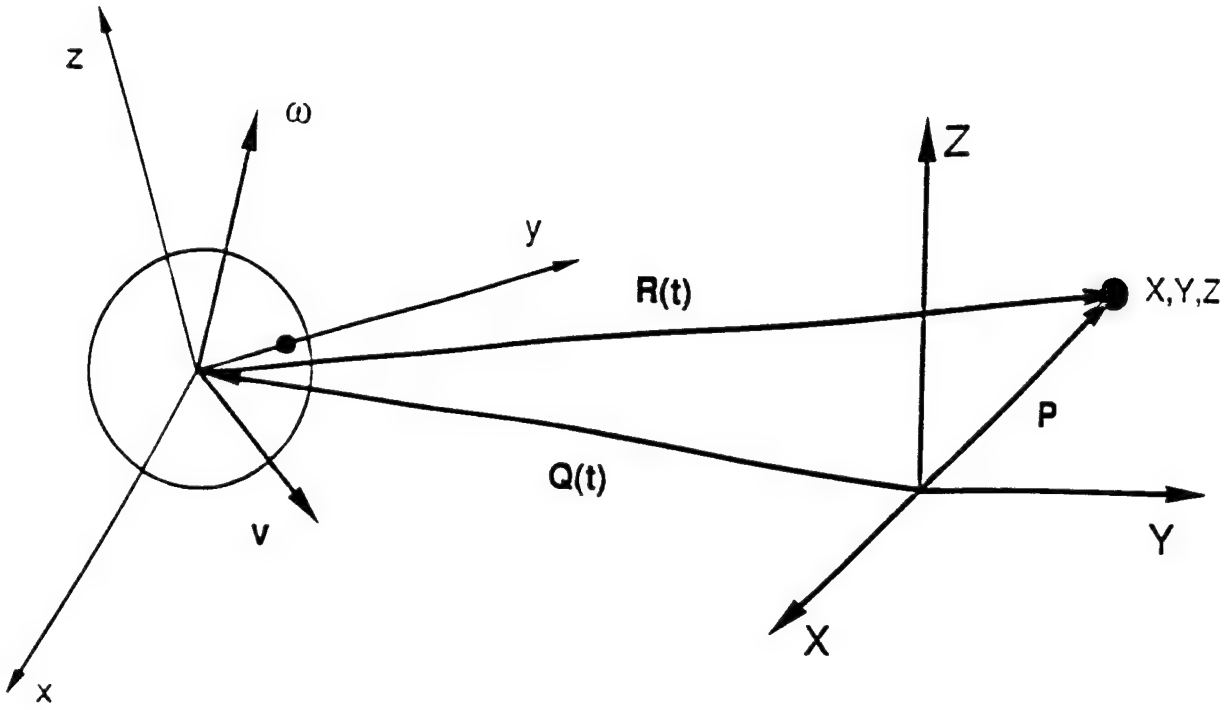


Fig. 1a. Vectorial relations between points in space and a moving eye given in world notation. **P** is the time invariant position of the point in object-centered system. **Q**(t) is a vector connecting between the systems origins. **R**(t) is a vector connecting the moving system origin and the point. The motion of the eye-system is depicted by the vectors **v** and ω given in eye-system notation.

The eye system is assumed to have 6 degrees of freedom. These are $\mathbf{v} = [v_x, v_y, v_z]$ and $\omega = [\omega_x, \omega_y, \omega_z]$, and denote the linear velocity and angular velocity vectors of the moving system, respectively. From fig. 1a it is clear that

$$\mathbf{P} = \mathbf{Q}(t) + \mathbf{R}(t) \ , \tag{1}$$

where $\mathbf{P} = [X, Y, Z]$ is the position vector of object point in the fixed system, $\mathbf{Q}(t) = [X(t), Y(t), Z(t)]$ is the vector connecting eye system and the fixed system origins. $\mathbf{R}(t)$ is a vector connecting the point and the eye system origin. Note that $\mathbf{P}$ is not time dependent.

The point $\mathbf{P}$ in the fixed system can be mapped into the eye system by

$$\mathbf{r}(t) = \underline{\mathbf{M}}(t)\,(\mathbf{P} - \mathbf{Q}(t)) \ , \tag{2}$$

where $\underline{\mathbf{M}}(t)$ is a 3x3 orthogonal transformation matrix ($\underline{\mathbf{M}}^T\underline{\mathbf{M}} = \underline{\mathbf{I}}$) the elements of which are functions of κ, λ, μ. The latter are taken here as Euler angles.

The absolute time derivative of $\mathbf{P}$ with respect to time is

$$\frac{d\mathbf{P}}{dt} = \mathbf{q}_t + \mathbf{r}_t + \omega \times (\mathbf{q} + \mathbf{r}) = \mathbf{0} \ , \tag{3}$$

Defining the eye system velocity as $\mathbf{v} = \mathbf{q}_t + \omega \times \mathbf{q}$ and substituting this expression in (3) yields

$$\mathbf{v} + \mathbf{r}_t + \omega \times \mathbf{r} = \mathbf{0} \ . \tag{4}$$

Each object point $\mathbf{r} = [x, y, z]$, given in terms of the eye system, is transformed into spherical retinal coordinate system (fig. 1b) by the following time invariant and one-to-one mapping.

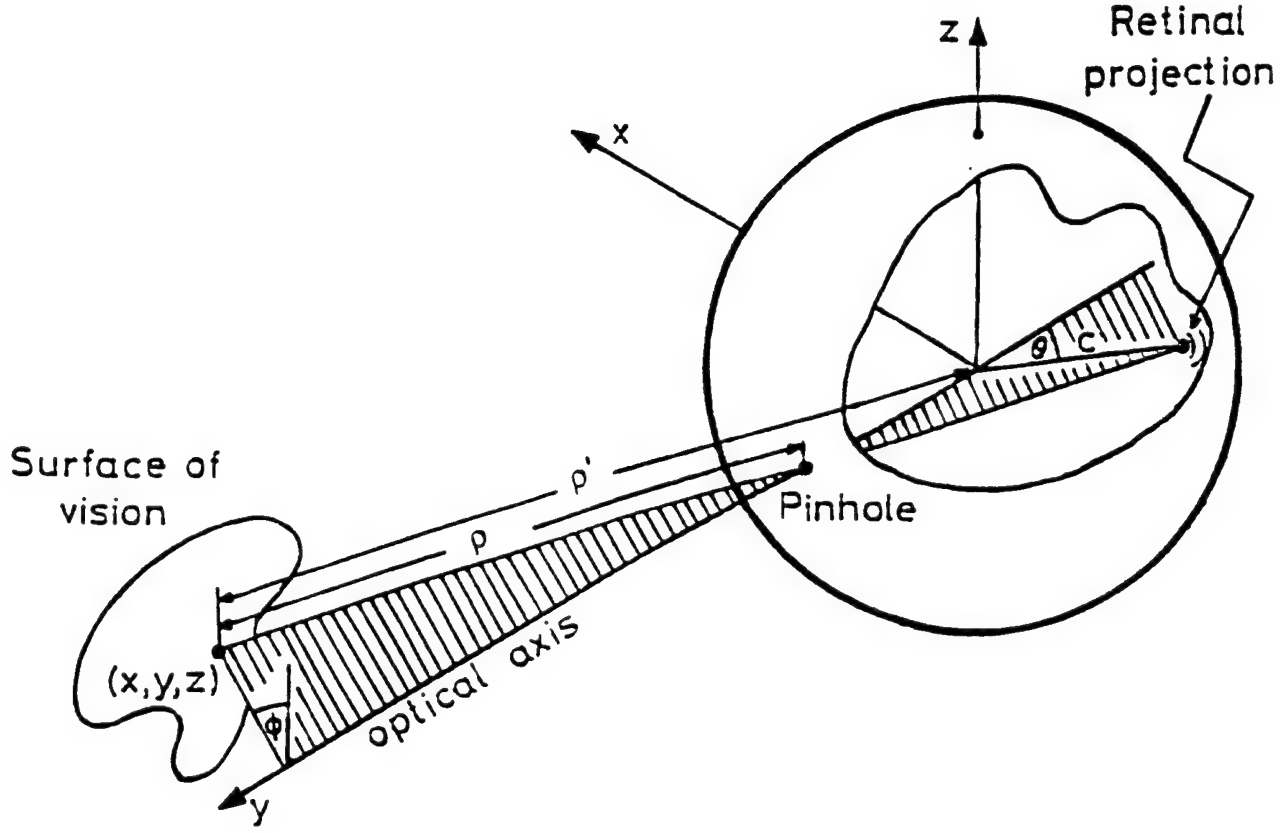


Fig. 1b. Definition of cartesian to polar mapping of surface points to retinal coordinates. θ and φ are eccentricity and meridian angles, respectively. ρ and ρ' are distances of a surface point from the pinhole and origin, respectively. c is the radius of the eyeball.

$$\theta = 2\ cos^{-1} \frac{y - c}{\sqrt{x^2 + (y-c)^2 + z^2}},$$

$$\varphi = tan^{-1} \frac{x}{z}, \qquad (5)$$

$$\rho = \sqrt{x^2 + (y-c)^2 + z^2},$$

$$\theta = 0 - \pi/2, \quad \varphi = 0 - 2\pi, \quad \rho = 0 - \infty$$

where θ, φ are retinal location and ρ is the unknown distance between the object point and the pinhole. The magnitude c representing the radius of the eyeball attains a significant role in the model by serving as a *metric unit.* The inverse mapping from retinal coordinates to eye system coordinates is

$$x = \rho\ sin \frac{\theta}{2} sin\ \varphi\ ,$$

$$y = c + \rho\ cos \frac{\theta}{2}, \qquad (6)$$

$$z = \rho sin \frac{\theta}{2} cos\ \varphi\ .$$

Differentiating (6) with respect to time and substituting in (4) yields three scalar differential equations in θ_t , φ_t, ρ_t. The magnitude ρ_t, which is not an observable quantity, can be

eliminated from this set of equations by sacrificing one, yielding the following flow equations:

$$\theta_t = \frac{2}{\rho}\left[v_y \sin\frac{\theta}{2} - (\rho + c\cos\frac{\theta}{2})(\omega_x \cos\varphi - \omega_z \sin\varphi) - \cos\frac{\theta}{2}(v_x \sin\varphi + v_z \cos\varphi)\right] , \tag{7}$$

$$\varphi_t = \frac{(\rho\cos\frac{\theta}{2} + c)(\omega_x \sin\varphi + \omega_z \cos\varphi) - (v_x \cos\varphi - v_z \sin\varphi)}{\rho \sin\frac{\theta}{2}} - \omega_y \quad . \tag{8}$$

The observable change of retinal representations of any point in the visual field is given by the two flow equations as a function of seven unknowns ρ, v_x, v_y, v_z, ω_x, ω_y, ω_z.

Since a point has, by assumption, no structure, ρ is taken as a scalar. Therefore, it follows that $\rho_\theta = \rho_\varphi = 0$. It turns out that a great simplification of the derivations is obtained by changing v_x, v_z and ω_x, ω_z into a new basis A, B, D, E as follows:

$$A = \omega_z \cos\varphi + \omega_x \sin\varphi , \tag{9}$$

$$B = \omega_z \sin\varphi - \omega_x \cos\varphi , \tag{10}$$

$$D = v_z \cos\varphi + v_x \sin\varphi , \tag{11}$$

$$E = v_z \sin\varphi - v_x \cos\varphi . \tag{12}$$

Substituting (9)-(12) in (7)-(8) and taking the first and second derivatives of the motion field with respect to θ and φ yields ten equations as shown in Appendix A. It turns out that not all of these

equations are independent. Moreover, their structure is such that all components of **v** are scaled by $1/\rho$. This is how the indeterminate scale problem is reflected in the present derivations. At this stage it is essential to show what information can be derived from these equations.

To isolate ω_y in Eq. (8), we take the second partial derivative of φ_t with respect to φ [or eq. (A10) in Appendix A] and find

$$\omega_y = -\varphi_t - \varphi_{t\varphi\varphi} \,. \tag{13}$$

A and B which are the equivalents of ω_x and ω_z have each several solutions, for example

$$A = \cot\frac{\theta}{2}\,\varphi_{t\varphi\varphi} - 2\varphi_{t\theta} = 2\,\frac{-\varphi_{t\varphi\varphi}\sin^2\frac{\theta}{2} + \theta_{t\theta\varphi}}{\sin\theta}\,, \tag{14}$$

$$B = 2\,\theta_{t\theta\theta} + \frac{\theta_t}{2} = \frac{-2\sin\frac{\theta}{2}\left(\varphi_{t\varphi}\sin^2\frac{\theta}{2} - \theta_{t\theta}\right) - \cos\frac{\theta}{2}\left(\theta_{t\varphi\varphi} + \theta_t\right)}{2\cos\frac{\theta}{2}\sin^2\frac{\theta}{2}}\,. \tag{15}$$

Note that equations (13)-(15) enable one to solve uniquely for $\boldsymbol{\omega}$ irrespective of ρ and **v**. This reduces the number of unknowns left to be found to four. Some independent relations can be derived for the other unknown magnitudes. First, two new variables are defined as

$$V_a = c\,A + E\,, \tag{16}$$

$$V_b = c\,B - D\,, \tag{17}$$

which enable one to solve for

$$\frac{V_a}{\rho} = -\frac{\theta_{t\theta\varphi}}{\sin\frac{\theta}{2}}, \tag{18}$$

$$\frac{V_b}{\rho} = -B\cos\frac{\theta}{2} - \varphi_{t\varphi}\sin\frac{\theta}{2}, \tag{19}$$

$$\frac{v_2}{\rho} = \frac{\theta_{t\varphi\varphi} + \theta_t}{2\sin\frac{\theta}{2}}. \tag{20}$$

Equations (18)-(20) comprise a set of three independent equations with four unknowns and cannot be solved uniquely. Note that V_a and V_b are both scaled by $1/\rho$. It can be shown that:

$$\Lambda = \frac{-\sin\phi + (V_b/V_a)\cos\phi}{-\cos\phi - (V_b/V_a)\sin\phi} = \frac{c\omega_x + v_z}{c\omega_z - v_x}, \tag{21}$$

where V_b/V_a can be computed as follows:

$$\frac{V_b}{V_a} = \frac{(2B + \theta_{t\phi\phi})}{2\,[\,A\cos(\theta/2) + \sin(\theta/2)\,\phi_{t\phi\phi}\,]\cos(\theta/2)}. \tag{22}$$

Thus, Λ is comprised of 4 pure motion parameters and is a position invariant constant for a given flow field.

D. A unique solution for two views

Given two views that enable extraction of all the aforementioned magnitudes, one can perform the following experiment: Find the predicted direction and magnitude of **v** by selecting arbitrary magnitudes of ρ. Figure 2 shows the result of such an experiment in which the predicted ρ of a given point is changed while $\boldsymbol{\Omega}$ is being kept fixed. The figure shows that the *direction* of the predicted **v** changes with changing ρ. This result is directly derived from Eqs. (18) and (19) in which V_a and V_b are compound variables of $\boldsymbol{\Omega}$ and **v** and scaled by $1/\rho$. Nonlinear relation between velocity and egodistance is contrasted with the linear dependence predicted by the indeterminate scale problem. Changes in the direction of **v** with different magnitudes of ρ indicates the existence of a unique solution to the problem, even for two views, because it means that there should be only one direction of **v** which is associated with veridical ρ.

At this stage of the analysis the navigational considerations enter. The navigational process is broken into a sequence of infinitesimally small discrete steps. Then, initial conditions have to be set. These are: an arbitrary starting orientation κ_0, λ_0, μ_0 required to set the value of $\underline{\mathbf{M}}(t_0)$, and an arbitrary origin of the world coordinates $\mathbf{q}(t_0)$. Note that setting the initial values does not involve selection of any of the unknown magnitudes. The state equation (1) is written for the first snapshot as

$$\mathbf{P} = \underline{\mathbf{M}}^{\mathrm{T}}(t_0)[\mathbf{r}(t_0) - \mathbf{q}(t_0)]. \tag{23}$$

Assuming that $\boldsymbol{\Omega}$ was calculated it is possible to calculate $\underline{\mathbf{M}}(t_1)$ and write the state equation for the second snapshot as

$$\mathbf{P} = \underline{\mathbf{M}}^{\mathrm{T}}(t_1)[\mathbf{r}(t_1) - \mathbf{q}(t_1)]. \tag{24}$$

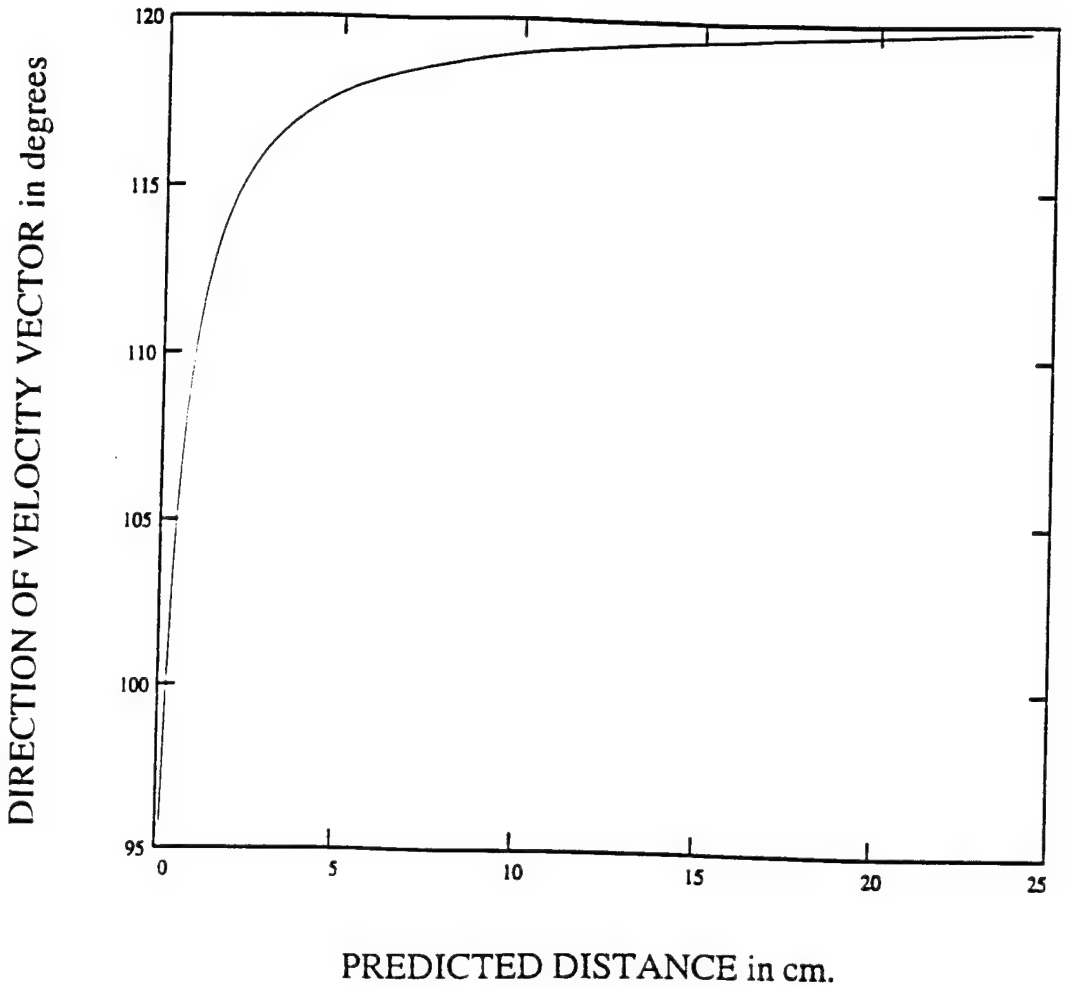


Fig. 2. The predicted direction angle between of the velocity vector **v** and ω is plotted against different values of predicted distance ρ. The true ρ is 12.1 cm. The non-linear relation is typical to the present definition of the optic flow and indicates the existence of unique solution for two views.

Since

$$\lim_{dt \to 0} \frac{\mathbf{Q}(t + dt) - \mathbf{Q}(t)}{dt} = \mathbf{Q}_t(t) = \mathbf{V}(t) \quad , \tag{25}$$

it follows

$$\mathbf{Q}(t_1) = \mathbf{Q}(t_0) + \mathbf{V}dt \quad . \tag{26}$$

Equation (24) can be written in eye system notation as follows

$$\underline{\mathbf{M}}^T(t_1)\mathbf{q}(t_1) = \underline{\mathbf{M}}^T(t_0)[\mathbf{q}(t_0) + \mathbf{v}dt] \quad . \tag{27}$$

Multiplying both sides of (25) by $\underline{\mathbf{M}}(t_1)$ yields

$$\mathbf{q}(t_1) = \underline{\mathbf{M}}(t_1)\underline{\mathbf{M}}^T(t_0)[\mathbf{q}(t_0) + \mathbf{v}dt] \quad . \tag{28}$$

Substituting $\mathbf{q}(t_1)$ in (24) and subtracting the resultant equation from (23) yields after algebraic manipulations

$$\underline{\mathbf{M}}^T(t_0)\mathbf{r}(t_0) - \underline{\mathbf{M}}^T(t_1)\mathbf{r}(t_1) + \underline{\mathbf{M}}^T(t_0)\mathbf{v}dt = \mathbf{0} \quad . \tag{29}$$

Multiplying both sides of (27) by $\underline{\mathbf{M}}(t_0)$ yields

$$\mathbf{r}(t_0) - \underline{\mathbf{M}}(t_0)\underline{\mathbf{M}}^T(t_1)\mathbf{r}(t_1) + \mathbf{v}dt = \mathbf{0} \quad . \tag{30}$$

Equation (30) comprises a set of 3 (linear) scalar equations. Taking dt = 1, these equations with any two of Eqs. (18)-(20) are written in discriminant form as a set of 5 linear equations with 5 unknowns as

$$\begin{bmatrix} m_{1,1} & m_{1,2} & 1 & 0 & 0 \\ m_{2,1} & m_{2,2} & 0 & 1 & 0 \\ m_{3,1} & m_{3,2} & 0 & 0 & 1 \\ m_{4,1} & 0 & 0 & 0 & -1 \\ m_{5,1} & 0 & 1 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \rho_0 \\ \rho_1 \\ v_x \\ v_y \\ v_z \end{bmatrix} = \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ c\omega_x \\ c\omega_z \end{bmatrix} \tag{31}$$

where the elements of the three first rows of (31) are

$$m_{i,j} = \begin{bmatrix} sin(\theta_0/\theta 2)\, sin\varphi_0 \\ cos(\theta_0/2) \\ sin(\theta_0/2)\, cos\varphi_0 \end{bmatrix}, \qquad i= 1, 2, 3 \quad j = 1 \ , \tag{31a}$$

and thus

$$m_{i,j} = -\underline{\mathbf{M}}(t_0)\underline{\mathbf{M}}^T(t_1) \cdot \begin{bmatrix} sin(\theta_1/2)\, sin\varphi_1 \\ cos(\theta_1/2) \\ sin(\theta_1/2)\, cos\varphi_1 \end{bmatrix}, \qquad i= 1, 2, 3 \quad j = 2 \ , \tag{31b}$$

and thus

$$k_i = c\, \underline{\mathbf{M}}(t_0)\, \underline{\mathbf{M}}^T(t_1) \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - c \cdot \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \ , \quad i = 1, 2, 3 \ . \tag{31c}$$

Note that the elements of $m_{i,1}$ and $m_{i,2}$ (with i = 1,2,3) are factored by ρ_o and ρ_1 respectively, because the free floating magnitude c becomes part of the constant element vector k_i. The last two rows of (31) are derived from eqs. (18) and (19) where

$$m_{i,j} = \begin{bmatrix} sin\varphi_0 & -cos\varphi_0 \\ cos\varphi_0 & sin\varphi_0 \end{bmatrix} \cdot \begin{bmatrix} -\theta_{t\theta\varphi 0}/sin(\theta_0/2) \\ -B_0 cos(\theta_0/2) - \varphi_{t\varphi 0} sin(\theta_0/2) \end{bmatrix}, \quad i = 4, 5 \quad j=1 \ . \tag{31d}$$

Note that *all* the components of the constant-elements-vector are scaled by c. The determinant of the coefficient matrix is

$$m_{1,1}m_{3,2} - m_{3,1}m_{1,2} - m_{1,4}m_{1,2} - m_{1,5}m_{3,2} \ . \tag{31e}$$

Substituting for $m_{i,j}$ and collecting terms the following relations have to hold for singularity of the solution

$$[sin\frac{\theta_0}{2} sin\varphi_0 - \frac{(c\omega_z - v_x)}{\rho_0}]\, m_{3,2} = [sin\frac{\theta_0}{2} cos\varphi_0 + \frac{(c\omega_x + v_z)}{\rho_0}]\, m_{1,2} \quad . \qquad (31f)$$

The actual solution to the problem, starting with eq. (23), seems to be difficult to follow and needs some clarifications. First it was based on the fact that $\boldsymbol{\omega}$ could be recovered for any view irrespective of $\mathbf{v}$ and ρ. This enabled to calculate the transformation matrixes $\underline{\mathbf{M}}_i$ Second, triangulation considerations were made that provided the 3 first rows of system (31). The triangle embodied in eq. (30) is shown in figure 3. It is created by the visual point, and the origins of the eye system at t_0 and at t_1. The figure depicts with solid lines the sides of this triangle where $\mathbf{R}_0$, $\mathbf{R}_1$, $\mathbf{V}dt$ are given in terms of the world system.

Indeed all the pertinent magnitudes have to be brought into the same space and time frame which is in this case the eye system at t_0. $\mathbf{r}(t_0)$ creates no problem because it is already given in term of this frame of reference. $\mathbf{r}(t_1)$, however, is needed to be transformed twice; first by applying $\underline{\mathbf{M}}^T(t_1)$ which transforms it into world system notation at t_1, then by applying $\underline{\mathbf{M}}(t_0)$ which transforms the latter into eye-system notation at t_0. For each of the triangle sides; $\mathbf{R}_0$ and $\mathbf{R}_1$, the computational system has its polar 3-D representation given in eqs. (6) with ρ_i as unknown. The last two rows of (31) deal only with the first view (t_0). They are derived from eqs. (18) (19) by multiplying both equations with ρ_0. In addition, V_a and V_b are transformed to their old basis by applying eqs. (9)-(12), and (16)-(17). We note that the same solution can be obtained by the differential approach without the need for the assumption about the point spread function and on the sole basis of the primitive flow equations (7)-(8). The latter solution is also based on the fact that $\boldsymbol{\omega}$ and all the pertinent magnitudes embodied in eqs. (18)-(19) can still be recovered irrespective of $\mathbf{v}$ and ρ when the scene is comprised of, at most, 5 rigid points (Kononov, 1996).

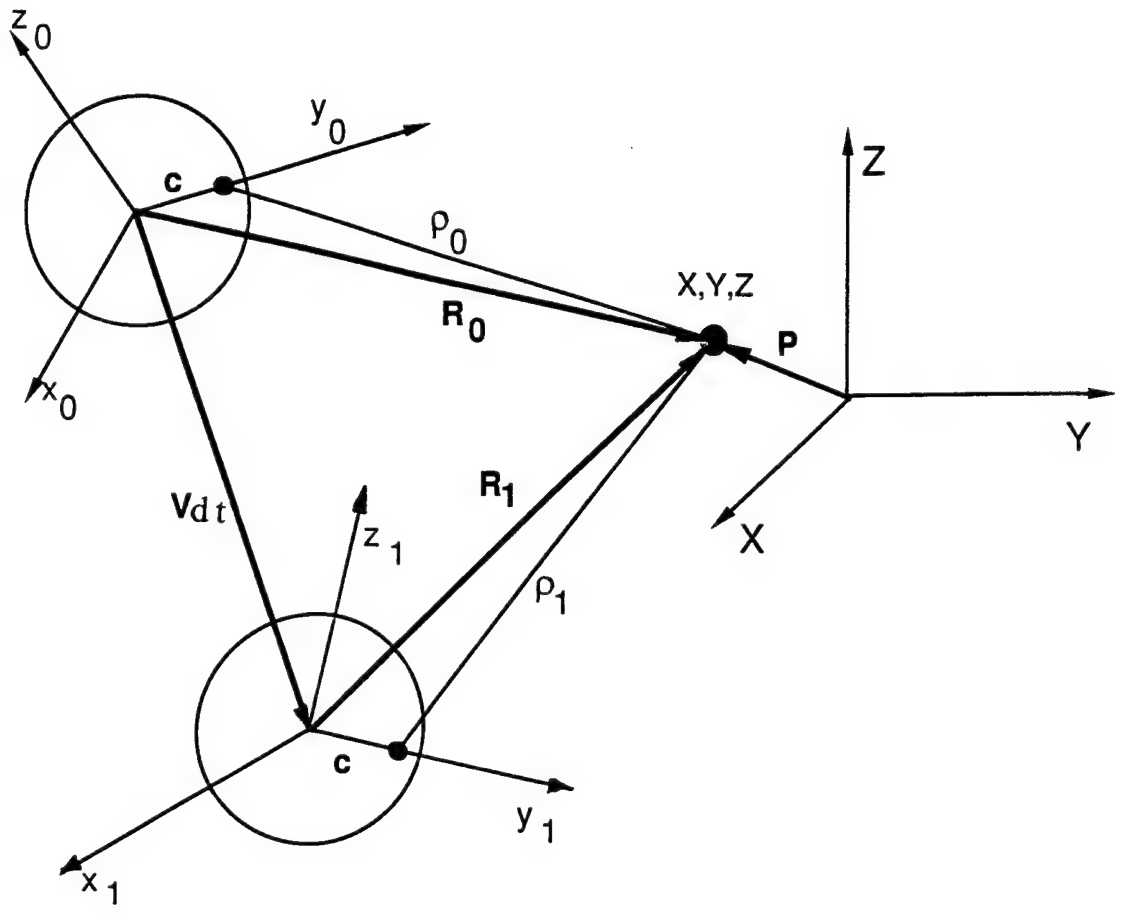


Fig. 3. The figure demonstrates how the solution for the navigational problem was obtained by triangulation. The two views of the world taken by a single eye are separated by the displacement **V**dt which establishes the base of the triangle. The basis and the other two sides of the triangle; $\mathbf{R}_0$ and $\mathbf{R}_1$, are transformed to their eye system notation at t_0, with ρ_0 and ρ_1 as unknowns This results in 3 independent equations where all notations are brought to a common frame of reference in time and space. In this scheme the selection of the world system origin is arbitrary.

A MathCad simulation of the algorithm was written and tested with many numerical examples. In general, the algorithm solved the navigational problem with an "infinite" precision. Moreover the three independent Eqs. (18)-(20) were found to be consistent in the sense that the use of any two of them produced an accurate solution, even for the same numerical examples. Inspection of the expression of the determinant of the coefficient matrix, and systematic investigation of the algorithm revealed several low probability singular conditions. These are

1) when the point lies on the optic axis,

2) when $m_{3,2}$ or $m_{1,2}$ equals zero. This requires that the pertinent entries of the matrix $\underline{\mathbf{M}}(t_0)$ $\underline{\mathbf{M}}^T(t_1)$ equals zero,

3) when $\omega = \mathbf{0}$. This is the most restrictive singular condition found.

A flow chart that summarizes the various stages of the algorithm is shown in fig. 4. The figure shows stages of the project-onto-the-retina and project-out-in-space of a static visual point. These involve the transformation of the point's spatial coordinates into retinal coordinates, followed by a stage of retinal motion extraction (see e.g. Hadani & Barta, 1989). The motion field is than processed to calculate the movement parameters of the eye in space and the distance-from-the-eye of that point. The motion parameters of the eye are then processed to calculate the project-

---

Insert Figure 4 here

Fig. 4. A flow chart depicting stages of the navigational model. These involve the transformation of the point's spatial coordinates into retinal coordinates (project-in), followed by a stage of retinal motion extraction. The motion field is then processed to calculate the egomotion parameters and the egodistance of the point. The egomotion parameters are processed to update the elements of the predicted inverse transformation matrix. The distance of the point with its retinal coordinates, accomplishes a 3-D retinal coordinate representation of the point. The 3-D retinal representation is projected-out-in-space at the output stage of the process. The figure also shows that extraretinal signals can be combined, by analog xor gates with retinal signals, when the latter are ineffective.

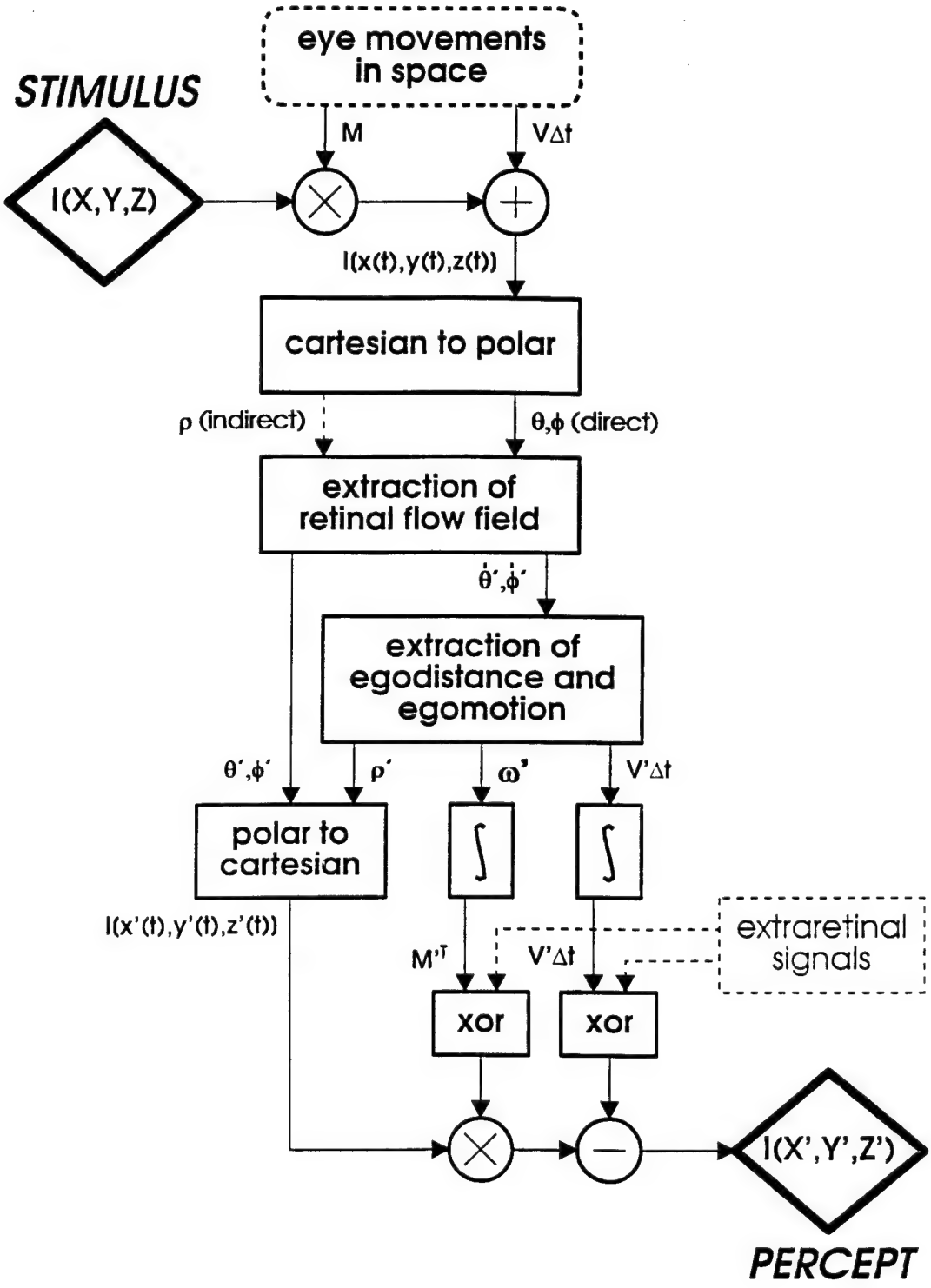

STIMULUS
eye movements in space
M
VΔt
I(X,Y,Z)
I(x(t),y(t),z(t))
cartesian to polar
ρ (indirect)
θ,ϕ (direct)
extraction of retinal flow field
θ̇′,ϕ̇′
extraction of egodistance and egomotion
θ′,ϕ′
ρ′
ω′
V′Δt
polar to cartesian
I(x′(t),y′(t),z′(t))
M′ᵀ
V′Δt
extraretinal signals
xor
xor
I(X′,Y′,Z′)
PERCEPT

Figure 4

out transform which is the inverse of the predicted project-in transform. The distance coordinate of the point with its retinal coordinates, accomplish a 3-D retinal coordinate representation of the point. The 3-D retinal representation is projected-out-in-space at the output stage of the process. The figure also suggest a way by which extraretinal signals can supplement retinal signals when the latter are ineffective (for example, during saccades).

2.3. DISCUSSION

In this work we show the feasibility of a scaled solution for the general passive navigation problem. Assuming that the retinal flow field is given, the investigation focused on the kinematical aspects of passive navigation. The solution obtained requires two perspective views and can be applied to the case when the input to the system is comprised of a single point. The first question to be discussed is what has made this solution possible?

The formalization of the problem presented here is different in several respects from those found in the common definitions of the optic flow (see e.g. Longuet-Higgins & Prazdny, 1980; Simpson, 1993). For example, the common definition position the origin of the moving system at the perspective center (nodal point, or pinhole), while the present definition positions it at the geometrical center of the eye. As a result, rotation of the eye produces a translation to the pinhole and makes the optic flow sensitive to distance due to eye rotations (see e.g. Hadani *et al.,* 1994). Second, in the present case any spatial point gets first its 3-D cartesian notation in the moving system and is then projected onto the retina with a one-to-one cartesian to polar transformation that preserves the distance as an unknown magnitude. This is different from the use of homogeneous coordinates in the common definition. Third, the single metric magnitude that exists in the system - the radius of the eyeball is not normalized to 1 as it is considered as an intrinsic scale. We note, however, that all of these differences together, or each by itself, could not produce the unique solution but turned out to be helpful and even crucial for the final solution. As long as the computations remained within the moving system notation, the problem was underconstrained.

The solution became unique only when the navigational procedures were applied. The latter required one to transform observable and unknown magnitudes into their object centered equivalents. Analysis of cause for lack a of unique solution in the eye system notation and the existence of a unique solution in object space revealed the following truism of calculus: The relations derived in the eye system were based on the differentiation of the position vector of the point in space ($d\mathbf{P}/dt$) as is done in Eq. (4). Applying the differentiation operation results in dropping out of the constant element in the fixed system leaving only the time varying elements. Since there is no time varying aspects of the position of rigid points within the object-centered system, the time derivative was justly set to zero. Navigation, on the other hand, is analogous to integration. In the case of differential equations, integration restores the constant element and requires one to set the initial conditions. In the present case it involved selection of world system origin ($\mathbf{q}_0$) and 3 orientation angles (κ_0, λ_0, μ_0). We argue that setting the initial conditions in the present case did not violate any strict requirement of passive navigation because it did not involve the need for external information. In conclusion, navigational considerations were essential to produce a solution to the passive navigation problem. The latter involved transforming eye-system magnitudes into object centered magnitudes. Projecting-out was aided by preserving the one-to-one mapping between object and retinal coordinates. Separation of the rotation center from the pinhole allowed the solution for two views. Scaling was obtained by considering the radius of the eyeball as an intrinsic metric unit.

While the implications of the present results to machine vision cannot be overlooked, the rest of this section will concentrate on the relevance of our procedures to human vision. First, we stress that the deterministic noiseless scheme presented above should be considered just as an ideal observer. Second, the motivation behind the development of this ideal observer model was to elucidate certain abstract theoretical ideas about the nature of mental representation as explained next.

One major difficulty for any theory of vision is that we are unaware of the existence of our eyes and seem to look through a hole in front of our head (see e.g. van de Grind, 1990, p. 392). This idea was nicely illustrated by the "Visual Ego" metaphor that was originally advanced by the physicist and psychologist E. Mach in 1906 (1959) and modified by Gibson (1979) as shown in fig. 5. The three drawings show different snapshots of Gibson's study as seen by his mind's eye while sitting on a comfortable chair. Each snapshot relates to a different heading direction. They demonstrate, in the first place, the idea about a window in the head through which the mind's eye observes the world. Second, they demonstrate the out-thereness of the perceptual experience, e.g. that objects are perceived out there in space and not as being displayed on some internal 2-D neural screen (Braunstein, 1976).

However, the three snapshots may not reflect directly one additional aspect of Mach-Gibson illustration. Let us suggest our own interpretation. The reader may note that Gibson's nose is shown in all three drawings occupying the same position. Gibson's nose is an object that is attached to his head, and in this respect it is an egocentered object. A TV set is shown in all three drawings and is one static external object that occupies different positions in Gibson's visual field. Now, the question is: which of the two, the TV set or the nose, has time invariant representation in Gibson's mind's eye? You can test and answer this question yourself by rotating your head about a vertical axis while fixating on a static object and attending to your nose (or any object that is firmly attached to your head). You may realize that external static objects are usually perceived as static while the nose is perceived as moving. You can also realize that by focusing attention on your nose you can perceive the nose as static and the external objects as moving. This "experiment" demonstrates that with attentional shifts one can change his/her static frame of reference.

The subjective impression demonstrated by the "Visual Ego" metaphor indicates that some low level unconscious processing is carried-out on the retinal optic flow before it become cognitively available. Had the mind's eye seen its retinal image directly, an instable inverted distorted and blurred scene will be seen as predicted from geometry and optics. Thus,

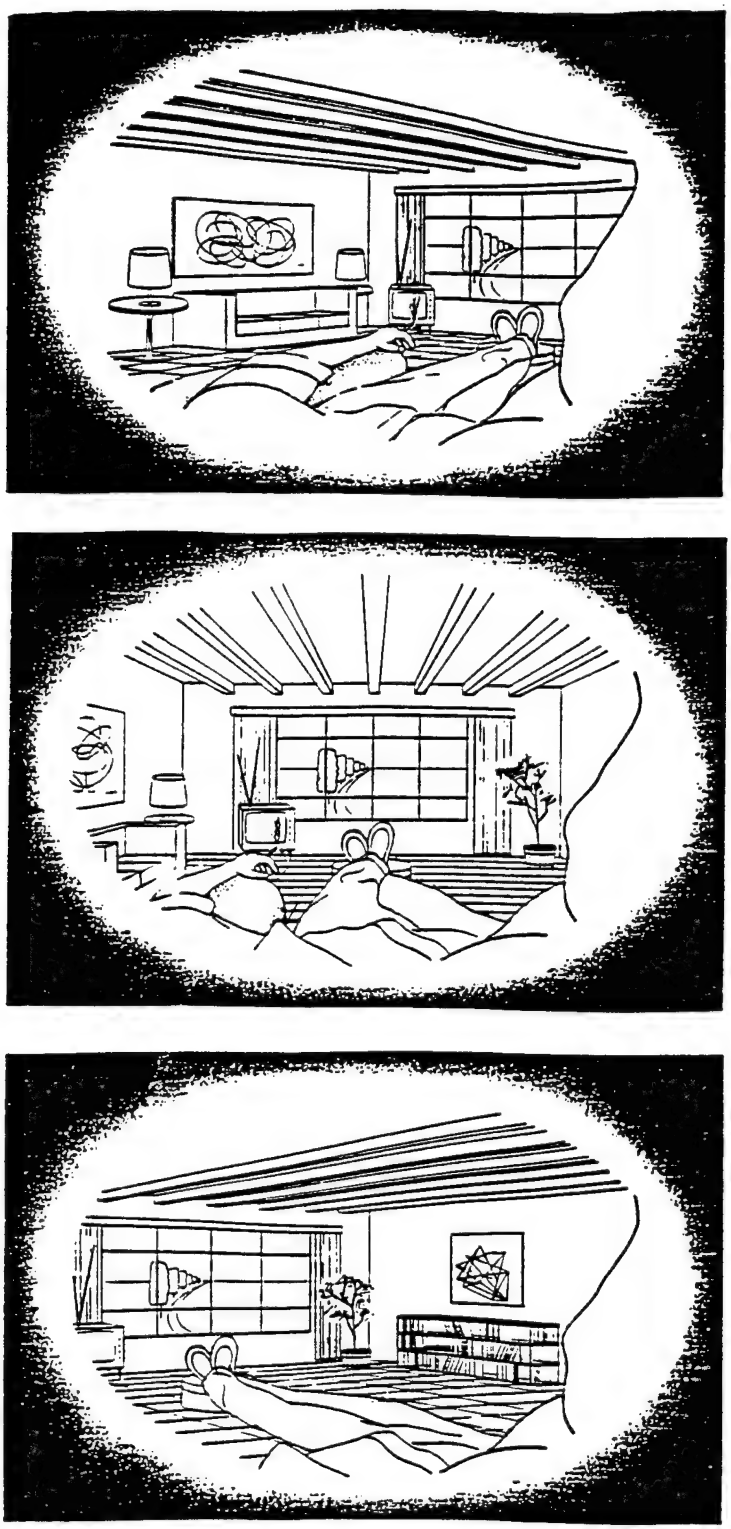


Fig. 5. Gibson's improvisation of the "Visual Ego". The three drawings show different snapshots of Gibson's study as seen by his mind's eye while sitting on a comfortable chair. Note that Gibson's nose, which is an egocentered object, occupies the same position in the three drawings. The TV set is an external object, but occupies different positions. The question is: which of the two, the TV set or the nose, has time invariant representation in Gibson's mind's eye? (from Gibson, 1979.)

computational vision models of structure from motion, and the SPIN theory in particular, may be considered as an attempt to bridge the gap between features of the retinal projection and our introspective impression. Alternative analyses have not arrived at a unique scaled solution. By showing here the feasibility of scaled solution, the theoretical difficulty created by the unscaled solutions is removed, thus paving the way to apply the navigational approach as a model for human vision as explained next.

Consider again the situation where a monocular observer is moving relative to a single static object. Then, for two successive views a retinal flow is created which is a pure effect of eye movements. By assumption, this flow is not visible but is presumably processed to recover egodistance and eye motion parameters. The latter magnitudes are then used to calculate the position of the object in object-centered coordinates. Our introspective impression indicates that only at this stage of processing the object becomes visible. Thus, according to the model, for these two successive views, the global slip of the retinal image is effectively filtered-out and the differential-motion parallax-component is used to extract the distance of the object. The output of this stage is non-inverted representation (coordinates) of the object out-there in space. Thus, the navigational approach accounts for visual stability, depth perception and the phenomenal experience of out-thereness. As a by-product, it also accounts for reinverting of the retinal image - a problem that according to Bridgeman *et al.* (1994, p. 257) was never solved. Let us assume that the representation in the first two views is kept in some spatiotopically mapped storage, and consider the third view (or any t+dt view). The same processing is applied now on the information in the second and the third views but with one difference. There is no need to redefine the initial conditions but just to update the elements of the coordinate transformation. This means that for the third view, the perceptual system may use *the same frame of reference* defined already for the first view. Eventually the coordinates of the object's points that are visible in all three views remain time invariant preserving position size and shape. In conclusion, the coordinate transformation model embedded in the navigational approach may provide an appealing theoretical framework to account

for three major aspects of visual perception: visual stability, object constancies and depth through motion parallax.

ACKNOWLEDGEMENTS

The authors are indebted to Dr. Gideon Ishai for fruitful discussions at various stages of this work and for the derivation of the basic kinematical equations, to Dr. Bela Julesz for his encouragement, support and reviewing the manuscript, to Dr. Stanley Klein for helpful comments and discussions, and to Ms. Carol A. Esso for her competent word processing. This work was supported by the U.S. Air Force Office of Scientific Research under grant 93-NL-165 to the first author. HLF was supported by NSF grant DMR 9023541.

REFERENCES

Braunstein, M. (1976). *Depth perception through motion.* New York: Academic Press.

Bridgeman, B., van der Heijden, A. H. C., & Velichkovsky, B. M. (1994). A theory of visual stability across saccadic eye movements. *Behavioral and Brain Sciences,* **17**, 247-292.

Bruss, A. R. & Horn, B. K. P. (1983). Passive navigation. *Computer Vision, Graphics, and Image Processing,* **21**, 3-20.

Cutting, J. E. (1986). *Perception with an eye for motion.* Cambridge: MIT Press.

Duric, Z., Rosenfeld, A., & Davis, L. S. (1995). Egomotion analysis based on the Frenet-Serret motion model. *International Journal of Computer Vision,* **15**, 105-122.

Fermuller, C. & Aloimonos, Y. (1995). Qualitative egomotion. *International Journal of Computer Vision,* **15**, 7-29.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Grind, van de, W. A. (1990). Smart mechanisms for the visual evaluation and control of self-motion. In R. Warren & A. H. Wertheim, (Eds.), *Perception and control of self-motion.* Hillsdale: Lawrence Erlbaum.

Hadani, I., Ishai, G., & Gur, M. (1978). Visual stability and space perception in monocular vision: mathematical model. *Technion-IIT Report # TSI 07-78.* Haifa, Israel: The Julius Silver Institute of Biomedical Engineering.

Hadani, I., Ishai, G., & Gur, M. (1980). Visual stability and space perception in monocular vision: mathematical model. *Journal of the Optical Society of America A,* **1**, 60-65.

Hadani, I., Kononov, A., Ishai, G., & Frisch, H. L. (1994). Two metric solutions to three-dimensional reconstruction for an eye in pure rotations. *Journal of the Optical Society of America A,* **11**, 5, 1564-1574.

Hadani, I. & Barta, E. (1989). The hybrid constraint equation for motion extraction. *Image and Vision Computing,* **7**, 3, 217-224.

Hadani, I. (1995, in press). The SPIN theory - a navigational approach to space perception. *Journal of Vestibular Research*, **5**, 6.

Hecht, S. & Mintz, E. U. (1936). The visibility of single lines at various illuminations and retinal basis of visual resolution. *Journal of General Physiology,* **22**, 593-612.

Henn, V., Cohen, B., & Young, L. R. (1980). Visual-vestibular interaction in motion perception and the generation of nystagmus. *Neurosciences Research Program Bulletin,* MIT Press, **18**, 4.

Horn, B. K. P. (1986). *Robot Vision.* Toronto: McGraw-Hill.

Howard, I. P. (1986). The vestibular system. In K. R. Boff, L. Kaufman, & J. P. Thomas, (Eds.), *Handbook of perception and human performance Vol. 1 - Sensory processes and perception* (11-1 to 11-30). New York: John Wiley and Sons, Inc.

Koenderink, J. J. & van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A*, **8**, 2, 377-385.

Kononov, A. (1996). The SPIN theory and the indeterminate scale problem. Master's thesis. New Brunswick, NJ: Rutgers University.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defence of weak fusion. *Vision Research,* **35**, 3, 389-412.

Longuet-Higgins, H. C. & Prazdny, K. (1980). The interpretation of moving retinal images. *Proceedings of the Royal Society of London B,* **208**, 385-387.

Longuet-Higgins, H. C. (1981). A computer algorithm for reconstruction a scene from two projections. *Nature,* **293**, 133-135.

Mach, E. (1959). *The analysis of sensations.* New York: Dover Publications, Inc.

MacKay, D. M. (1973). Visual stability and voluntary eye movements. In R. Jung (Ed.). *Handbook of sensory physiology* (**7**, 3A). New York: Springer.

Meiri, A. Z. (1980). On monocular perception of 3-D moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **PAMI-2**, 582-583.

Nagel, H. H. (1981). Representation of moving rigid objects based on visual observations. *Computer,* **14**, 8, 29-39.

Prazdny, K. (1981). Determining the instantaneous direction of motion from optical flow generated by a curvilinear moving observer. *Computer Graphics and Image Processing,* **17**, 238-248.

Simpson, W. A. (1993). Optic flow and depth perception. *Spatial Vision,* **7**, 1, 35-75.

Tsai, R. Y. & Huang, T. S. (1985). Uniqueness and estimation of 3-D motion parameters and surface structures of rigid objects. In W. Richards & S. Ullman (Eds.), *Image Understanding 1985-86* (Chap. 6). Norwood, NJ: Albex.

Ullman, S. (1979). *The interpretation of visual motion.* Cambridge: MIT Press.

APPENDIX A

List of first and second partial spatial derivatives of the optic flow equations. These are obtained after substituting the new variables A-E in Eqs. (7), (8).

$$\theta_{t\theta} = -\frac{\sin\frac{\theta}{2}}{\rho}(cB - D) + \frac{v_y}{\rho}\cos\frac{\theta}{2}\ , \tag{A1}$$

$$\theta_{t\varphi} = 2A + \frac{2(cA + E)}{\rho}\cos\frac{\theta}{2}\ , \tag{A2}$$

$$\theta_{t\theta\varphi} = \theta_{t\varphi\theta} = -\frac{cA + E}{\rho}\sin\frac{\theta}{2}\ , \tag{A3}$$

$$\theta_{t\theta\theta} = -\frac{1}{2\rho}\left[\cos\frac{\theta}{2}(cB - D) + v_y\sin\frac{\theta}{2}\right]\ , \tag{A4}$$

$$\theta_{t\varphi\varphi} = -2B - 2\frac{cB - D}{\rho}\cos\frac{\theta}{2}\ , \tag{A5}$$

$$\varphi_{t\theta} = -\frac{\rho A + \cos\frac{\theta}{2}(cA + E)}{2\rho\sin^2\frac{\theta}{2}}\ , \tag{A6}$$

$$\varphi_{t\varphi} = -\frac{\rho B\cos\frac{\theta}{2} + (cB - D)}{\rho\sin\frac{\theta}{2}}\ , \tag{A7}$$

$$\varphi_{t\theta\theta} = \frac{2\rho A\cos\frac{\theta}{2} + (cA + E)(1+\cos^2\frac{\theta}{2})}{4\rho\sin^3\frac{\theta}{2}}\ , \tag{A8}$$

2

$$\varphi_{t\theta\varphi} = \varphi_{t\varphi\theta} = \frac{\rho B + \cos\frac{\theta}{2}\,(cB - D)}{2\rho \sin^2\frac{\theta}{2}}, \qquad \text{(A9)}$$

$$\varphi_{t\varphi\varphi} = -\frac{\rho A \cos\frac{\theta}{2} + (cA + E)}{\rho \sin\frac{\theta}{2}}, \qquad \text{(A10)}$$

CHAPTER 3:
THE EFFECT OF INTER-OCULAR DISTANCE ON REGISTERED DEPTH IN RDS WITH DIFFERENT PEDESTAL DISPARITIES

1. INTRODUCTION

Stereoscopic depth constancy has been a subject of research and speculations for more then a century, yet it is not fully understood. The literature is large and diverse (see Ono and Comerford 1977; Foley 1991; Durgin et al. 1995, for reviews), and the constancy hypothesis still has proponents and opponents indicating that the problem is a difficult one. The present study focuses on one narrow aspect of this issue, i.e. the question of how the metric measure of interocular distance (IOD) does affect depth appreciation in random dot stereogram (RDS). Geometry of binocular vision shows that disparity is scaled by the inverse of the square of the viewing distance. This is what is known as the inverse square law (ISL) which also sets the relation between the latter magnitudes and depth and IOD (see Appendix). While there are many studies with conflicting results on the effects of viewing distance and disparity on perceived depth, little work has been done on the effect of IOD. The latter effect can be investigated either by optical modification of the natural IOD (Wallach et al. 1963; Wallach and Karsh 1963; Fox and Mauk 1991), or by using individual differences in IOD as is done in the present study. The evaluation of the effect of this metric measure has theoretical and practical implications.

As an anthropometric measure the IOD varies considerably among individuals. In adult females it ranges from 4.6 cm (5th percentile) to 7.46 cm (95th percentile) (Woodson et al.1992). In adult males the corresponding values are 5.7 cm and 6.96 cm. Thus, the total range is over 2.86 cm which is large enough to produce significant differences in perceived depth. [With the use of young observers the lower limit of this range can be extended even further]. In many studies, however, the IOD was taken as one of the random variables and eventually it contributed to the error variance. In the present study we utilized the IOD as a quasi-independent variable to test whether its effect on perceived depth is indeed as predicted by the ISL (see Fig. 1).

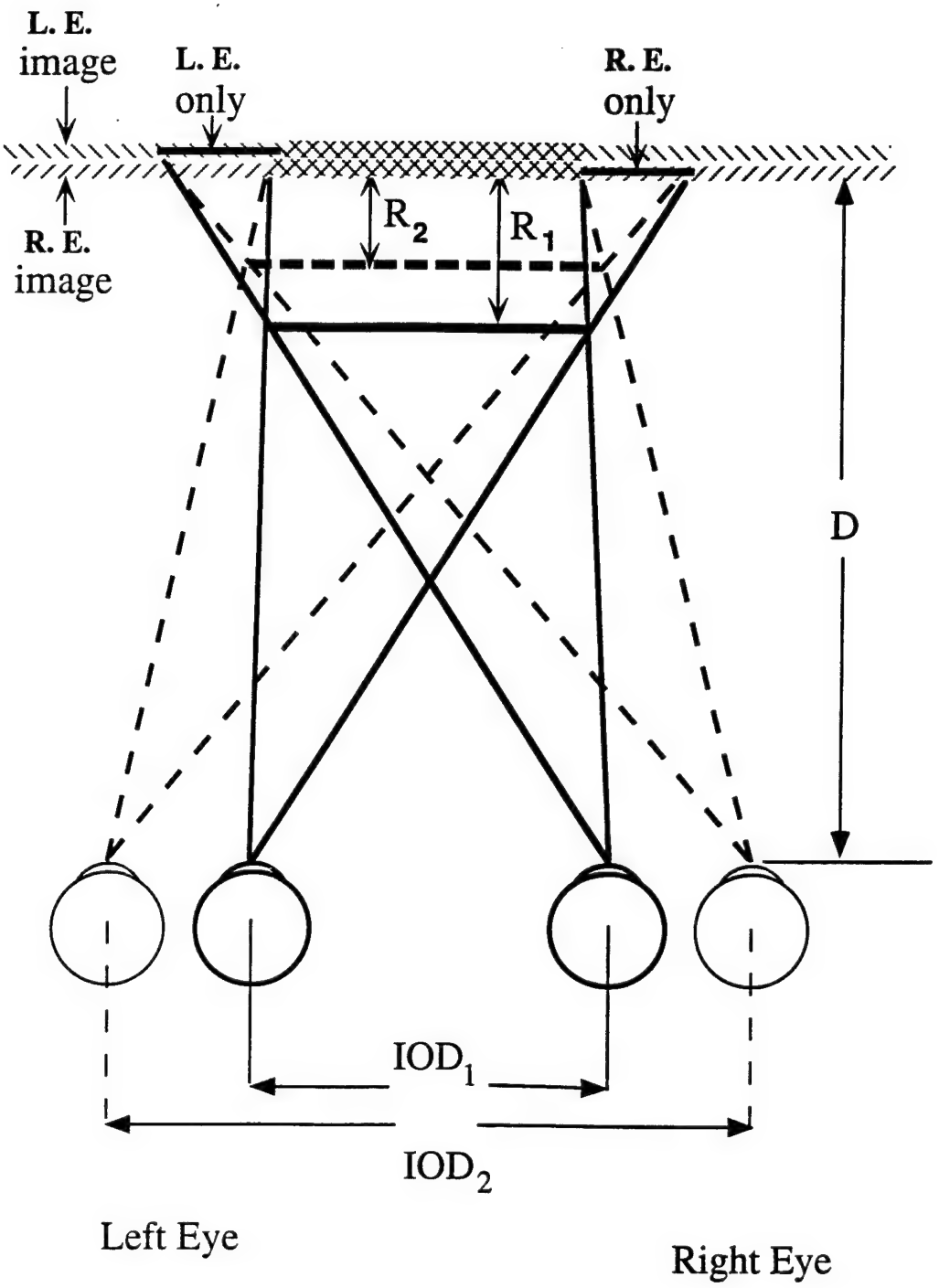


Figure 1. Geometry of stereopsis in viewing RDS in anaglyph form. The two half images are depicted by hatched lines. The figure depicts the perceived depth of the hovering square by two observers viewing the same RDS from the same viewing distance. Note that for observer 2, the hovering height is smaller and the size of the square is larger.

A manifestation of the effect of IOD is obtained with the use of telestereoscope (Wallach et al. 1963; Wallach and Karsh 1963; Fisher and Ebenholtz 1986). The simplest telestereoscope is a modified Wheatstone stereoscope having two pairs of mirrors (ocular and objective pairs) instead of one. This arrangement alters the effective IOD and the optical viewing distance. As a result, a rotating real and rigid object appears deformed (Wallach et al. 1963). Several studies investigated the effect of adaptation to the deformation and its aftereffect (Wallach et al.1963; Wallach and Karsh 1963; Fisher and Ebenholtz 1986). These studies did not involve the quantification of the primary effect of the telestereoscope but provided evidence that the device modifies perceived depth of real objects. Fox and Mauk (1991) utilized the telestereoscope to modify the effective IOD and viewing distance to a stereogram. In a factorial 3x3 within subject design (3 IODs and three viewing distances), the effects of the two variables were found to be significant. Their results were also in line with the predications of the ISL but with systematic depth underestimations. This type of study cannot tell whether individual differences in IOD evoke differences in depth appreciations of different observers. In conclusion, there is little but unambiguous evidence that alteration of IOD within observers modifies stereoscopic depth for both real and simulated disparity. Thus, while individual differences in IOD do not seem to alter depth appreciations of real objects, the question of whether the same differences alter depth appreciations in stereograms remains open.

The psychophysical task utilized in this study involved a moveable probe that had to be aligned independently to the same depth of the central square and of the background of a stereogram. This task was considered as being inadequate to making measurements which may project on the issue of depth constancy because it does not involve the subjective evaluation of depth interval. The task is named "disparity matching" (Johnston 1991) because it can be performed by equating (annulling) the disparities generated by the probe or the target and does not convey information about perceived depth of the probe or the target. Thus, in one of the present correlational studies we minimized the possibility of disparity matching, i.e. the possibility that the subjects utilized a magnitude other then depth to make their judgments. We also utilize here the

notion of registered rather then perceived depth to account better for what was actually measured as a dependent variable. Moreover, being aware of the restrictions that this psychophysical method has in inferring perceived depth, our conclusions were derived accordingly (see discussion).

In a preliminary study, the stimulus was a RDS in anaglyph form that was figure 1 in Julesz and Miller (1975) paper. The central square had disparity of 5 arcmin (2 mm) which calculated to evoke depth between 2.3 and 3.3 cm for IODs of 7.3 and 5.0 cm respectively, at a viewing distance of 85 cm. However, a low and statistically insignificant effect of IOD was found. We conjectured that this low effect resulted from the relatively low magnitude of the disparity that the central square had, and because of the pedestal disparity that this stimulus had. To test these possibilities we conducted 4 correlational studies in which the square had higher disparity magnitudes and the background had zero, crossed and uncrossed disparities.

2. METHODS

Observers were presented RDS in anaglyph form. The stereogram portrayed a central square with a crossed disparity and occupied 1/3 of the background field. The experimental setup is shown in figure 2. In four correlational experiments, subjects had to align real probes superimposed by a beam splitter independently with the depth of the square and the background. The viewing distance was at 86 cm and was kept within +/- 0.5 cm error with the use of a chinrest. The experimental room was kept dark except for the lumination emanating from the CRT screen and from a 40 watt incandescent table lamp that illuminated the probe. Unless otherwise specified, the stereogram was comprised of 96x96 dot elements which occupied visual angle of 8.8^0x8.8^0. Each dot element was comprised of 2x2 red or green pixels of identical luminance. The dot element subtended 5.5 arcmin which also defined one disparity unit.

In the first two correlational studies the anaglyph was presented on a MAC II 21" color monitor and the background had zero disparity. In one of these studied the density of the RDS was 50% of bright dots and double real bars served as probes. In the second, the density was 5% of

bright dots and the probe was a single real point light source. The probes were hooked to a single 12" linear rack and pinion slide with a scale and were superimposed by a beam splitter on the observer's field of view. The double probe was comprised of two vertical bars lying on the same plane and painted in yellow to have them being seen binocularly through the red green glasses. One probe of 2x3 cm was superimposed just below the edge of the central square. The second probe 2x5 cm was superimposed 0.5° aside the left of the background. Observers had to align the short probe to be seen at the same depth of the square and to repeat the same procedure with the long probe for the background. The point light source was comprised of a flat polished yellow LED that was masked with 1 mm wide aperture. This probe was superimposed roughly at the center of the central square in an area free of bright dots. Observers had to align the depth of the probe to

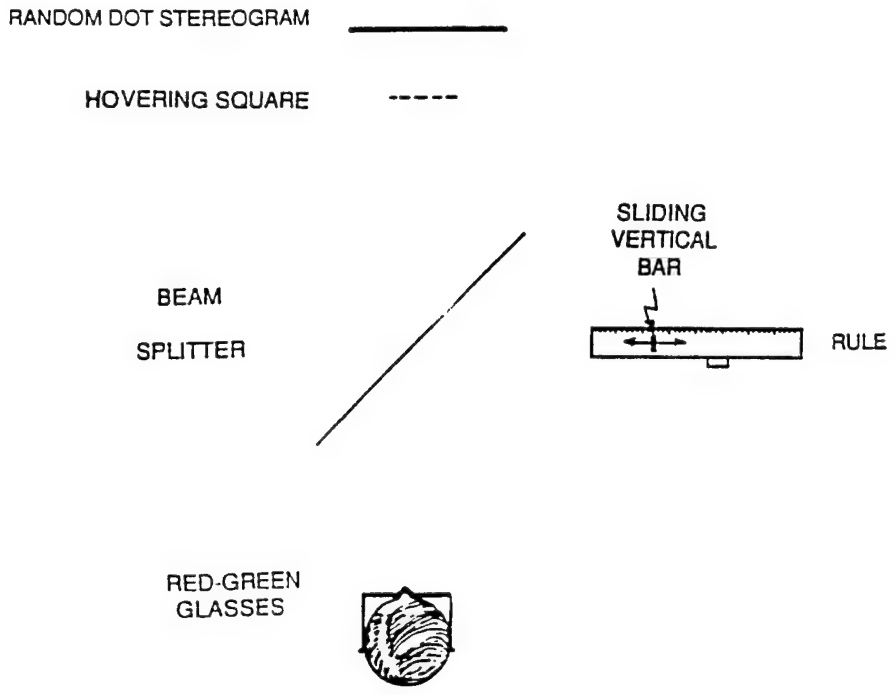


Figure 2. The experimental setup. RDS in anaglyph form was presented on a CRT screen. Subjects aligned a real probe seen superimposed on the RDS by the beam splitter. The probe was a dual vertical bars or a single point light source.

coincide with the depth of the square and then to make the same alignment of the probe with the background. This type of probe diminished judgments which are based on disparity matching, particularly for the background. The initial depth of the probes relative to the targets was varied at random. For each target the subjects repeated the task 6 times. Except for a few subjects no training trials were required.

The second two correlational experiments were replications of the former high density experiment with stereograms having crossed and uncrossed pedestal disparities. The methods were identical except for the following differences: In the uncrossed pedestal experiment, the size of image was 11.8°x11.8° comprised of 128x128 dot elements. The background had -2 disparity units and the square had +6 disparity units relative to the background. Thus the square had +4 disparity units relative to the display's surface. In the crossed pedestal experiment the background had +2 disparity units and the square had +4 disparity units relative to background. Thus, the square had totally +6 disparity units relative to the display surface. Moreover, in this experiment the anaglyph was presented on a paper print illuminated by a second table lamp. Thus, the display was viewed without the intervening effects of the CRT screen. The latter had a thick glass that elongated the optical distance and a curvature that made it act as a lens. Moreover, there was some minute pedestal disparity produced to the anaglyph because the red and green phosphor elements did not coincide on the CRT screen.

3. SUBJECTS

Each experiment was carried out on a large pool of observers who had good stereopsis. Stereopsis was tested by the ability to identify simple forms presented with anaglyphs. The ages ranged from 5 to 63 years old of different gender and race, with and without corrected vision. The interpupillary distance (IPD) of the observers was measured with corneal reflection pupillometer - Essilor model C16103 - for convergence at the viewing distance (86 cm) and for parallel convergence. The latter measure was taken as the observer's IOD. A total of 163 observers participated, about 25% served in more than one experiment.

4. ANALYSIS

Data of the 6 alignments made by each subject of registered depth of the square and the background were averaged to yield double scores.Then, averaged scores were plotted against IOD (and IPD) as scatter plots separately for the background and the square. A logarithmic curve-fit was calculated for each set of the data with Macintosh's Statwork™ software. This procedure also yields the Pearson correlation coefficient. In addition, the differences between these two measurements were computed. Then, the Pearson correlation coefficient was calculated between the logarithm of these differences and the logarithm of (IOD + Separation). The latter two magnitudes are predicted to have a linear relationship by the ISL (see Appendix A). Similarly, the Pearson correlation coefficient r was calculated between the logarithm of the differences and the logarithm of (IPD + Separation). The goodness of fit between the theoretical predictions and the data was evaluated by calculating the root mean square (RMS) of the data from predictions.

5. RESULTS

The registered depths of the square and background obtained in the four experiments plotted against IOD are given in figures 3 a-d. The dashed lines show the theoretical predictions. The respective Pearson correlation coefficients r are given in table 1 separately for IOD and IPD. Figs. 3 a and 3 b show the depth appreciations of 59 and 22 observers made in the high and low density experiments respectively. Both figures manifest a strong effect of IOD for the central square and no effect of IOD for the background. The values of r computed for the central square are -.892 ($p < 0.001$) for high density stimulus, and -.933 ($p < 0.001$) for the low density stimulus. The respective values for the background are -.141 (n.s.) and -.032 (n.s.). The corresponding values for the IPD are slightly lower but all are significant as well (table 1). Figs. 3 c and 3 d show the depth appreciations of 42 and 40 observers made in the crossed and uncrossed pedestal experiments respectively. The two figures manifest a strong effect of IOD for both the central square and the background.The values of r computed for the central square are -.760 ($p <$

0.001), and -.800 ($p < 0.001$) for the crossed and uncrossed pedestal disparity respectively. The respective values for the background are -.545 ($p < 0.001$) and -.596 ($p < 0.001$). The corresponding values of r computed for the IPD are lower for the crossed pedestal then those computed for the IOD, but are slightly higher for the uncrossed pedestal (table 1).

The depth intervals between the square and background calculated by subtracting the registered depth of the square from that of the background, plotted against IOD, are given in figures 4 a-d. The dashed lines show the theoretical predictions. The respective Pearson correlation coefficients r are given in table 2 separately for IOD and IPD. Figs. 4 a and b show the depth interval calculated for the high and low density experiments respectively. Figs. 4 c and d show the depth interval calculated for the crossed and uncrossed pedestal disparity experiments respectively. All four figures show an effect of IOD. The values of r are -.906 ($p < 0.001$) for high density stimulus, and -.953 ($p < 0.001$) for the low density stimulus. The respective values for crossed and uncrossed disparity stimuli are -.522 ($p < 0.015$) and -.815 ($p < 0.001$). The corresponding values for the IPD are slightly lower in two of the four experiments, almost even for the high density stimulus and lower for the crossed disparity stimulus (table 2).

		Square			
Experiment	n	r(iod)	p	r(ipd)	p
High density	59	-.892	.001	-.888	.001
Low density	22	-.933	.001	-.866	.001
crossed	40	-.760	.001	-.715	.001
uncrossed	42	-.800	.001	-.805	.001
		Background			
Experiment	n	r(iod)	p	r(ipd)	p
High density	59	-.141	n.s.	-.126	n.s.
Low density	22	-.032	n.s.	-.071	n.s.
crossed	40	-.545	.001	-.467	.005
uncrossed	42	-.596	.001	-.616	.001

Table 1. The Pearson correlation coefficients r computed for the four experiments between scores of registered depths of background and square against observers IOD and IPD. n. s. stands for not significant.

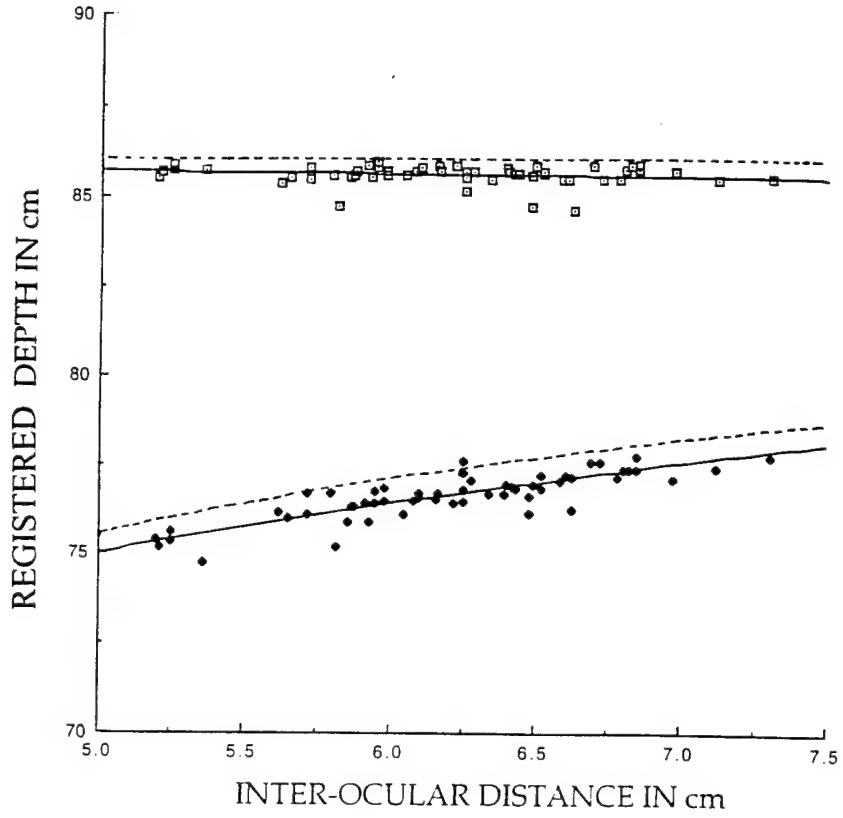
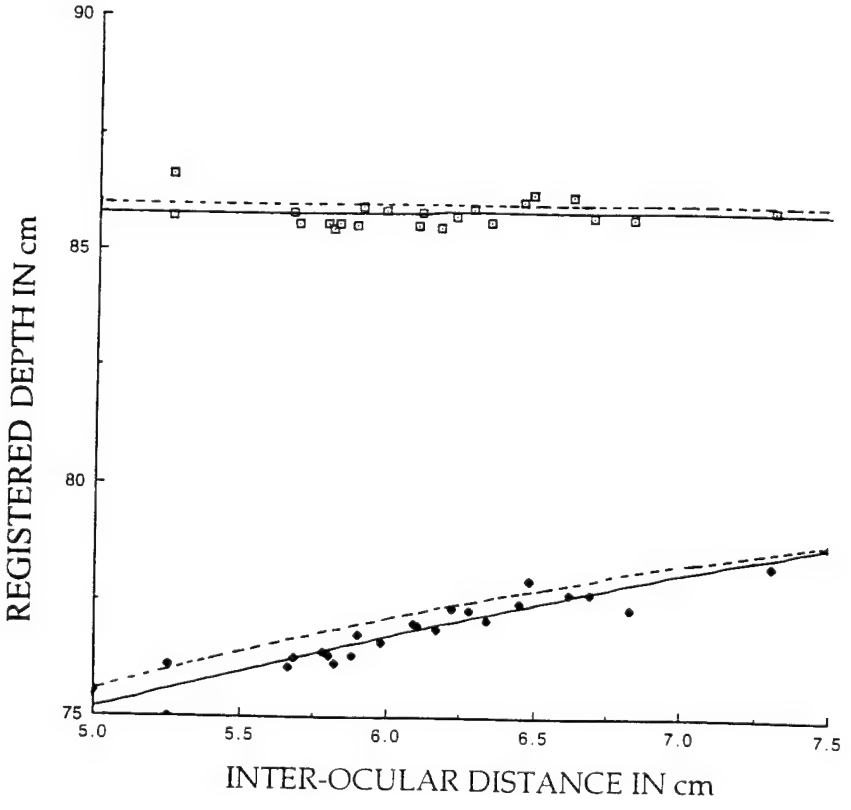


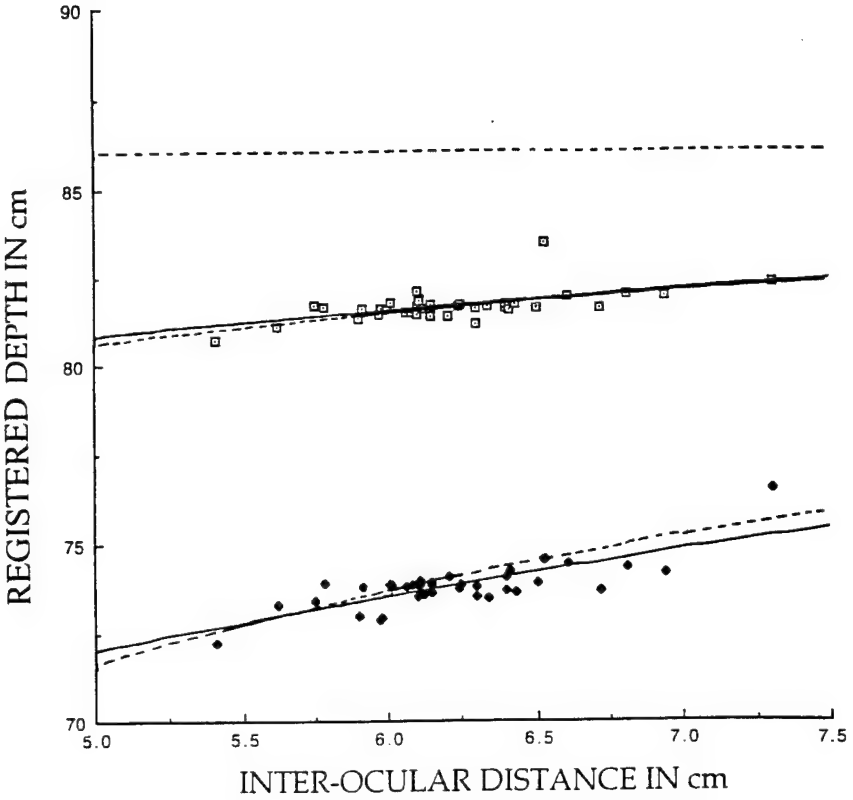
Figure 3 a. b.

Registered depth plotted against observers's IOD. Lower dashed line is the theoretical prediction. Upper dashed line at 86 cm indicates viewing distance, or distance to the display plane

a) Results of 59 subjects in the high density experiment. r = -.141 for background (upper curve), and r = -.892 for square (lower curve).



b) Results of 22 subjects in low density experiment. r =-.032 for background (upper curve) and r = -.933 for square (lower curve).



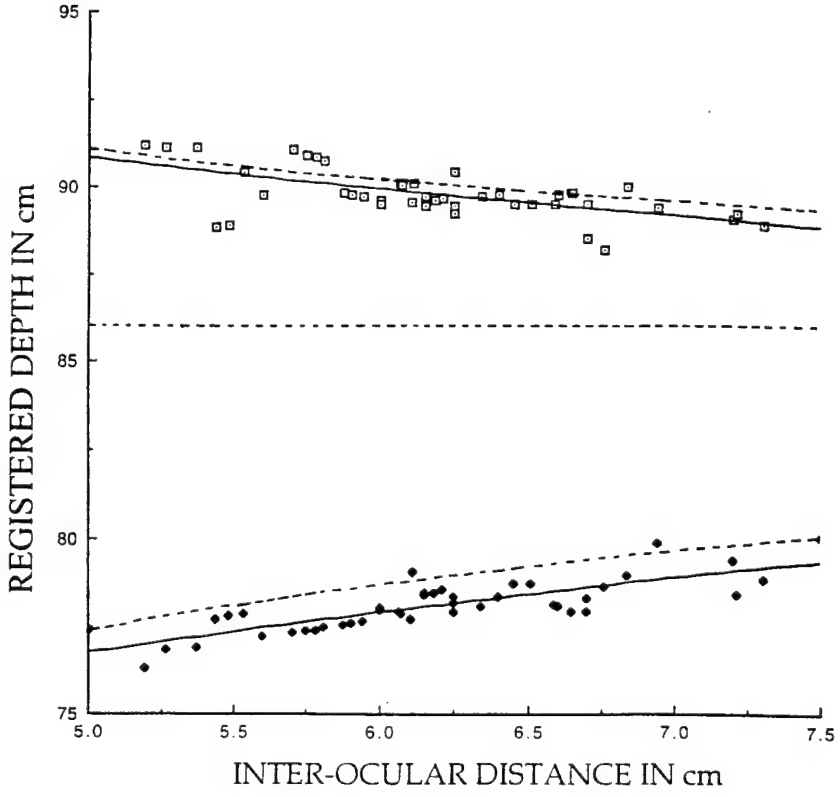


Figure 3 c, d.

Registered depth plotted against observers's IOD. Lower dashed line is the theoretical prediction. Upper dashed line at 86 cm indicates viewing distance, or distance to the display plane

c) Results of 40 subjects in the crossed pedestal experiment, r = -545 for background (upper curve), r = -.860 for square (lower curve)

d) Results of 42 subjects in uncrossed pedestal experiment r = -.595 for backgound (upper curve) and-.800 for square (lower curve).

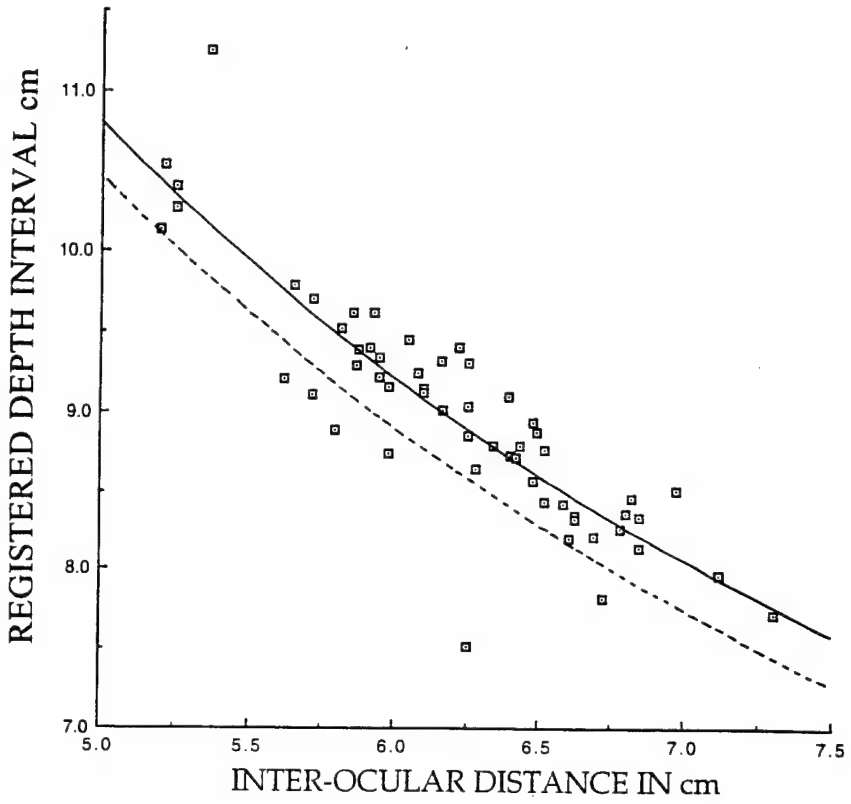
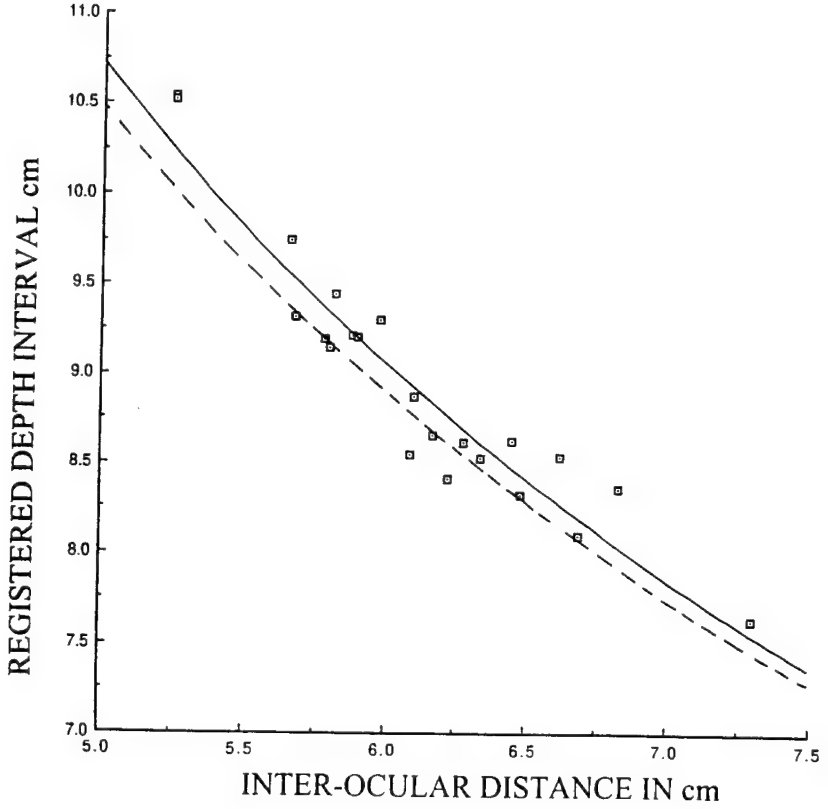


Figure 4 a, b.

Registered depth interval plotted against observers's IOD. Dashed line is the theoretical prediction.

a) Results of 59 subjects in the high density experiment r = -.906.



b) Results of 22 subjects in low density experiment r = -.953.

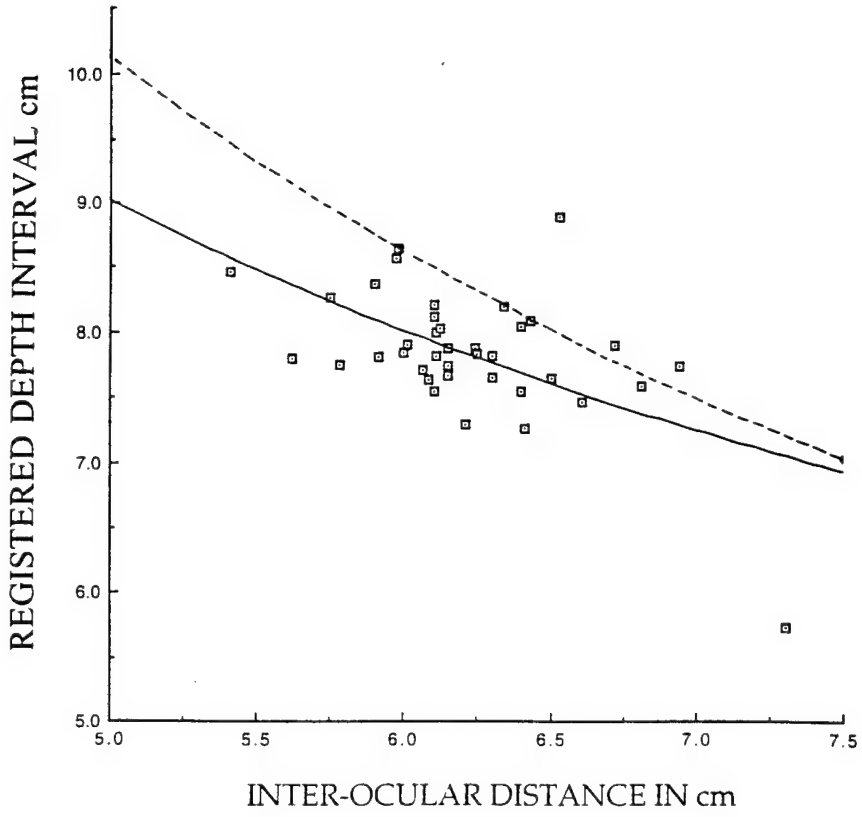
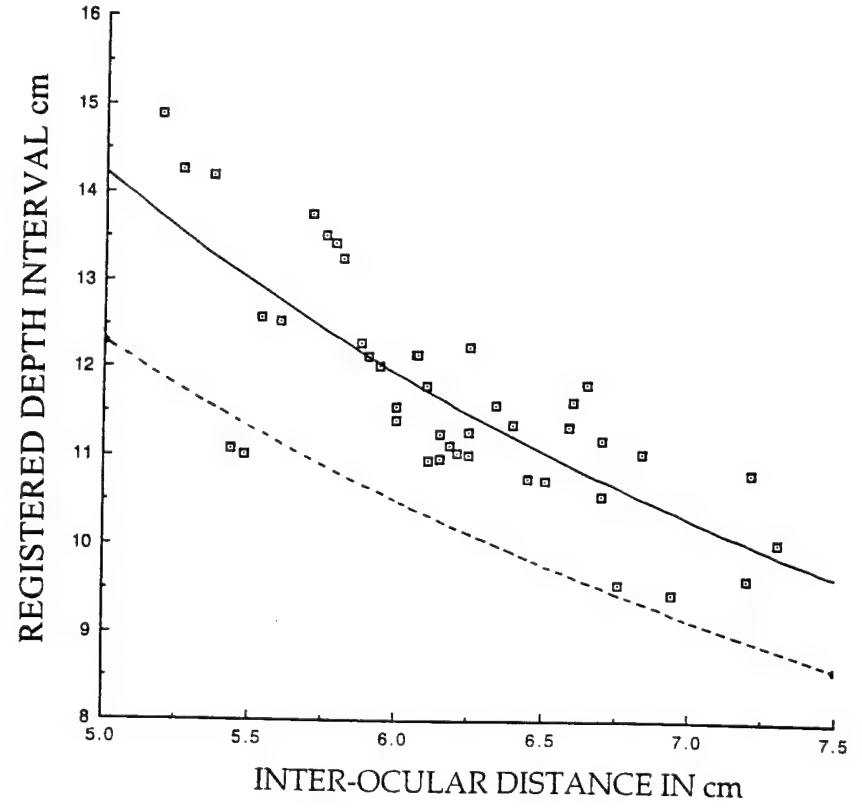


Figure 4 c, d.

Registered depth interval plotted against observers's IOD. Lower Dashed line is the theoretical prediction.

c) Results of 40 subjects in the crossed pedestal experiment r = -.522.



d) Results of 42 subjects in uncrossed pedestal experiment r = -.815.

		Registered depth interval			
Experiment	n	r (iod)	p	r (ipd)	p
High density	59	-.906	.001	-.908	.001
Low density	22	-.953	.001	-.938	.001
crossed	40	-.522	.001	-.544	.001
uncrossed	42	-.815	.001	-.769	.001

Table 2. The Pearson correlation coefficients r computed for the four experiments between scores of log (R) and log (I + S) when R is the registered depth interval and I is IOD or IPD.

Data for all experiments was compared to the theoretical predictions calculated from the ISL. The theoretical predictions are depicted in the scatter plots with dashed lines. The closest match between theoretical and empirical data was obtained in zero pedestal disparity experiments. Of these two stimuli the best match is shown for the low density case (figs. 3b and 4b). The goodness of fit in all experiments is given in table 3 in cm by calculating the RMS values of the data from the theoretical curve. The disparities between the theoretical and empirical curves is larger in the pedestal disparity experiments as compared to those obtained in the no pedestal disparity experiments.

Experiment	RMS square	RMS background	RMS differences
High density	.588	.248	.211
Low density	.448	.281	.226
crossed	.322	.472	.668
uncrossed	.636	.878	1.627

Table 3 Root mean square in cm of empirical data from the theoretical predictions for the 4 experiments measures in cm the goodness of fit. The best fit is obtained for the low density stimulus.

6. DISCUSSION

In general, our data show a clear effect of esoteric IOD on the registered depth evoked by simulated disparity. In the first place, this effect suggests that different observers viewing the same stereogram from the same viewing distance would experience quantitatively different depths. These quantitative differences are indicated by trends in the measured registered depth. Observers with smaller IODs had seen the square at closer egodistance then observers with larger IODs (Fig. 1 and figs. 3a and 3b). This effect is not a negligible one, for example, the depth interval in the low density experiment registered by observer with IOD of 7.3 cm was 7.7 cm while observer with the lowest IOD registered an interval of 10.6 cm (Fig. 4b). Moreover, the data shows: a) High correlations of up to -.953 between the registered depth of the central square and IOD. b) High correlations of up to -933 for the depth interval (as calculated by the subtracting the magnitudes of two independent measurements), and IOD. c) The fit of data with the theoretical predictions made by the ISL given in terms of RMS which was as low as 0.211 mm, and without the need of using a free parameter (Table 3 and dashed lines in fig. 4a). Thus, the present data provides additional empirical support for ISL which makes it valid to predict individual differences in registered depth. It is interesting to note that when comparing between the fits of the two pedestal experiments to their theoretical predictions, the fit of the crossed pedestal data was better in terms of RMS. This is probably due to the fact that the crossed pedestal stereogram was presented on a paper print, thus removing some of the biases in depth estimates introduced by the CRT screen.

Indeed, the fit of data with the ISL depended on stimulus conditions. The best fit with predictions was obtained in the high and low density experiments where the background had zero disparity. Zero disparity for the background also means that its depth is not simulated but coincides with the depth of the display's surface, or the horopter. This also means that matching between a real probe and between the background is identical to matching depth between two real objects. That is presumably why in this case there was no effect of IOD for the background (top of figs. 3a and 3b) and the correlations were lower and insignificant (table 1). On the other hand, when the

background had a crossed (uncrossed) disparity, the registered depth measurements indicated that it was perceived in front of (behind) the display plane. This is presumably because disparity of the background was taken by the observer's stereoscopic mechanism as a simulated disparity and the effect of the IOD is shown also on its registered depth (figs. 3c and 3d). In conclusion, our results suggest that three different surfaces have to be considered in viewing a stereogram: The surface of the figure, the surface of background, and the surface of the display. The data also suggest that under all the present stimulus conditions, the surface of the display served as the horopter. Since the ISL tacitly assumes depth constancy, then our data suggest that with zero pedestal disparity for the background depth is veridical, and that veridicality is somewhat impaired when the background has pedestal disparity. It is also interesting to note that in most cases, the IOD gave better overall predictions than IPD. This finding is mostly manifested in the zero pedestal experiment and deserves further investigation.

The results of the experiments with crossed and uncrossed pedestal solved what we called the pedestal dilemma, e.g. why in the preliminary study a low and insignificant correlation between IOD and depth interval was found. The pedestal dilemma is explained by the increased variance in depth matching for stereogram with pedestal disparity. Since the measurements for the square and the background were independent the variance of the differences is added and masked the effect of IOD. Then higher magnitudes of disparities of the square were needed for the effect to be revealed. A different question is raised by the data as for the possible cause(s) of the increased variance in the pedestal experiments. Even though this question was not addressed here, it is likely that it was caused by: a) The increased workload imposed by stereograms with pedestal disparity on the convergence and the accommodation mechanisms (Diner and Fender 1993). b) Visual deficiencies that observers may have that determined the strength of their stereopsis. If we take the ISL as a norm for good stereopsis, and the task of matching depth between real and simulated disparity as an accurate and sensitive one, then the question of how different visual deficiencies are reflected in the performance of this task is interesting and deserves further research.

It has been justly argued that results based on the task of disparity matching does not convey information about perceived depth of the probe or the target (Johnston 1991). Against this criticism, we argue first, that the task does tell that the two matched stimuli occupy the same location in space, or what we call here the registered depth. The differences between the two independent measurements, however, was not assessed directly by the subjects and may not provide meaningful information about the perceived depth interval but only if the observers did not change their horopter in switching between the two parts of the task. Theoretically, changing the horopter modifies the sign and the magnitude of the percieved depth. The good match, however, between the theoretical predictions and the registered depth interval, particularly in the no pedestal disparity experiments, suggests that the observers utilized the display surface as the horopter for both measurements (see below). Third, even though one may have a metric scale of depth, direct assessment of depth interval may suffer from other problems, for example, difficulties in mentally transforming 3-D perceptions into 2-D side view responses (Durgin et al. 1995). In contrast, the present method gives consistent results, and is easily communicable to a large population of observers including young children. Fourth, in some stereoscopic depth constancy studies, disparity matching was found to be an appropriate task for inferring veridicality of depth perception (Patterson et al. 1992). In other study it was found to be consistent with alternative measures of depth interval but was more accurate (Cormack 1984). In addition, the task has the benefits of the high resolution of the stereoscopic mechanism that falls in the hyperacuity range (2-5 arcsec). Finally, the criticism implies that observers may annul a magnitude other than perceived depth in their judgments. This possibility was diminished in the low density experiment where the probe was a single central point-light-source and was utilized for depth matching with both the square and the background.

For several reasons it is not obvious that the ISL is an accurate model for stereoscopic depth perception, even for real objects. Thus, any empirical support for law is theoretically important. The ISL is derived on the basis of strict Euclidean geometry premises and on the

principle of triangulation. In the derivation of the law, the visual axes of the two eyes are assumed to fixate steadily and intersect at a point. The fixation point and the two nodal points of the eyes create the Veith-Muller circle which defines the horopter, or the reference surface of zero disparity. However, horizontal phoria will cause the optical axes to intersect at a point different from the fixation target and the slightest vertical phoria would rule out any intersection. Furthermore, in normal vision the visual axes of the eyes are not steadily maintained, even when the eyes fixate on a steady target, because there are involuntary movements of fixation (Dichburn 1973), some of which may not binocularly synchronized (see Alpern 1972). Julesz and Fender (1967) have shown that under stabilized image conditions the two overlapping and fused half images can be pulled apart horizontally as much as 2° without the loss of fusion and/or stereopsis. Replicating this experiment under normal viewing conditions, Hyson et al. (1983) have shown that permanent fluctuations of vergence angle occur with no loss of fusion or stereopsis for vergence error of up to 5°. These findings indicte that convergence during fixation is not time invariant quantity and had disparity been scaled by convergence angle, as suggested by Cumming et al. (1991), fluctuations in perceived depth of fixated object are to be expected. These findings also suggest that a more advanced dynamic computational model of stereopsis is required in which the IOD, being roughly a time invariant base of the triangle, is the scaling magnitude. Moreover, even the static definition of the horopter assumes that there is, about any such circle, a range of disparities that can be resolved without the loss of single vision. This automatically produces for any given horopter certain thickness in the z dimension. The thickness of any horopter depends on a range of disparities that determine the region of single vision called the Panum fusional area (about 10 arcmin for brief exposure and 2° for longer exposure). Thus, the depth intervals calculated here from the difference of the two registered depth measurements may be considered as perceived depth intervals only if the two measurements fall within the same horopter.

Recent works on stereoscopic depth constancy which utilized real objects have shown good depth constancy (Durgin et al. 1995; Frisby et al. 1996). On the other hand, works utilizing

stereograms, gave conflicting results. For example, Johnston (1991) and Tittle et al. (1995) reported on systematic distortions in perceived shape of a semi-circular cylinder, and Norman et al. (1996) reported on a failure of depth constancy in the assessments stereoscopic oblique bars. In contrast, Durgin et al. (1995) reported on good depth constancy also for a stereoscopic cones. It is possible that at least some of the systematic distortions of shape reported by Johnston (1991) and Tittle et al. (1995) were caused by the use of mirrors' telestereoscope that eventually modified the natural IOD. But the effect was not taken into account in the calculations of depth (M. Landy, personal communication 1996). On the other hand the telestereoscope utilized by Durgin et al. (1995) was comprised of two prisms inclined in 45° which were placed vertically and may seem to unchange the natural IOD. Thus, this configuration may account for the reported depth constancy obtained in this study. In conclusion, our findings suggest that the natural or modified IOD should be measured and taken into account in psychophysical studies of stereoscopic depth constancy.

There is a wide consensus in vision that disparity is an ambiguous cue for depth because the same magnitude of disparity may evoke different depths depending on the viewing distance. The question is how the viewing distance is extracted stereoscopically? The ambiguity of disparity as a cue for depth stems from a fundamental computational problem associated with the horopter model which suggests that the viewing distance cannot be uniquely extracted by a mechanism that measures only angles. Therefore, the stereoscopic mechanism is considered as providing only relative depth and this feature cannot account for perceptual constancy. On the other hand, Mayhew and Longuet-Higgins (1982) suggested that vertical disparity enables the system to calculate absolute depth from retinal information. This conjecture is disputable and had received only a weak empirical support. Recently, Rogers and Bradshaw (1993) have shown that vertical disparity is an effective cue but only when the visual field is large and subtends 70°. The use of stereograms subtending 30^0, however, did not show an effect of vertical disparity (Cumming et al. 1991). In our view, there is no computational need to resort to vertical disparity for extracting absolute depth. In chapter 2 of this report, a model is presented that shows that, in principle, absolute depth can be

extracted from two monocular views of a single eye. This model utilizes the time domain for the computations. Since depth from motion problem is isomorphous with depth from stereopsis, the model in chapter 2 can be elaborated to account for binocular vision too. In the elaborated model, the IOD would serve as an intrinsic scale as suggested by the SPIN theory (Hadani et al.1980; Hadani 1991; Hadani 1994) similar to the radius of the eyeball that serves as an intrinsic scale for monocular vision. Such extended model has the benefit of allowing asynchronous eye movements (see also Hadani 1980), and may solve for the inadequacies of the classical horopter model mentioned above.

It may be questionable whether our findings have any theoretical significance regarding the plausibility of veridical depth perception in humans, but they certainly have practical implications. In the first place, they imply that depth in stereoscopic displays, like 3-D movies, will be scaled differently by different observers even if their viewing distance from the display is identical. Second, they are related to head (helmet) mounted displays used to fly aircraft (Patterson et al. 1992). These displays intend to provide the pilot a field of view under low visibility conditions (night vision). In these displays, the images are picked up by two cameras. If one takes the nodal points of the cameras' objective-lens as the vantage points, then the distance between the two nodal points is analogous to the natural IOD. And if this distance is larger than the operators' IOD, then hyperstereo information is created. Our findings suggest that in order to provide to users veridical depth, the distance between the cameras should be customized. [Another requirement, that falls beyond the scope of this work, is that the cameras' nodal points should coincide effectively with the natural nodal points (Hadani 1991)]. The same arguments apply to virtual reality displays that are becoming popular and widespread. In this case a synthetic environment is computer generated in a method called volumetric ray-tracing. A 3-D scene is embedded in a volume filled with texture elements. The back plane of the volume is the image plane. To generate each eye's image, a line is traced from the eye's view through the surface of the scene to an image pixel. Many of the virtual reality systems combine visual and motor (kinesthetic) displays that permit telerobotic manipulation

(Diner and Fender 1993). Thus in order to make the visual input compatible with the kinesthetic and reduce workload, the volumetric ray-tracing procedure should consider also the particular operators' IOD.

ACKNOWLEDGEMENTS

Two brief accounts of this work were presented: one in ARVO 1992 meeting (zero pedestal experiments), and the second in ARVO 1995 meeting (pedestal experiments). The authors wish to thank Aravind K. Suri for his help in data collection and analysis, and to Carol A. Esso for editing. The project was sponsored in part by AFOSR Grant 93-NL-165.

REFERENCES

Alpern, M. (1972). Eye Movements. In Handbook of Sensory Physiology. D. Jamson (ed.). Springer: Berlin VII/4, 303-330.

Cormack, R. H. (1984). Stereoscopic depth perception at far viewing distances. Perception and Psychophysics, 35, 5, 423-428.

Cormack, R. and Fox, R. (1985). The computation of disparity and depth in stereograms. Perception & Psychophysics, 38, 4, 375-380.

Cumming, B. G., Johnston, E. B. and Parker, A. J. (1991). Vertical diaparities and perception of three-dimensional shape. Nature, 349, 411-413.

Diner, D. B. and Fender, D. H. (1993). Human Engineering in Stereoscopic viewing Devices. Plenum Press: New York.

Ditchburn, R. W. (1973). Eye Movements and Visual Perception. Clarendon: Oxford.

Durgin, F. H., Proffitt, D. R., Olson, T. J., and Reinke, K. S. (1995). Comparing depth from motion with depth from binocular disparity. Journal of Experimental Psychology, Human perception and Performance, 21, 3, 679-699.

Fisher, S. K. and Ebenholtz, S. M. (1986). Does perceptual adaptation to telestereoscopically enhanced depth depend on the recalibration of binocular disparity? Perception and Psychophysics, 40, 101-109.

Foley, J. (1991). Binocular space perception. In D. Regan (ed.) Vision and visual disfunction, Vol. 9 Binocular vision. Boca Raton, FL: CRC Press, 75-92.

Fox, R., and Mauk, D. L. (1991). Increases in IPD reduces perceived depth in stereograms. 32nd Annual Meeting of the Psychonomic Society, San Francisco, 127, p. 12.

Frisby, J. J., Buckley, D., and Duke, P. A. (1996). Evidence for good recovery of lengths of real objects seen with natural stereo viewing. Perception, 25, 129-154.

Graham, C. H. (1965). Visual space perception. In C. H. Graham (ed.), Vision and visual perception. New York: Wiley.

Hadani, I., Ishai, G., Gur, M. (1980). Visual stability and space perception in monocular vision: mathematical model. J. Opt. Soc. Am., 1, 60, 60-65.

Hadani, I. (1991). The corneal lens goggle and visual space perception. Applied Optics, 38, 28, 4136-4147.

Hadani, I. (1994). The SPIN theory- a navigational approach to space perception. J. of Vestibular Investigation, 5, 6, 443-454.

Hyson, T. M., Julesz, B., and Fender, D. H. (1983). Eye movements and neural remapping during fusion of misaligned random-dot stereograms. J. Opt. Soc. Am., 73, 1665-1673.

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. Vision Research, 31, 7/8, 1351-1360.

Julesz, B., and Fender, D. H. (1967). Extension of Panum fusional area in binocular stabilized vision. J. Opt. Soc. Am., 57, 819-830.

Julesz, B. and Miller, J. E. (1975). Independent spatial-frequency-tuned channels in binocular fusion and rivalry. Perception, 4, 125-143.

Mayhew, J. E. W. and Longuet-Higgins, H. C. (1982). A computational model of binocular depth perception. Nature, 297, 376-378.

Norman, J. F., Todd, J. T., Perotti, V. J. and Tittle, J. S. (1996). The visual perception of three-dimensional length. J. of Experimental Psychology: Human Perception and Performance, 22, 1, 173-186.

Ono, H. and Comerford, J. (1977). Stereoscopic Depth constancy. In Stability and Constancy in Visual Perception. W. Epstein (ed.). John Wiley: New York.

Patterson, R., Moe, l. and Hewitt, T. (1992). Factors that affect depth perception in stereoscopic displays. Human Factors, 34, 6, 655-667.

Rogers, B. J., and Bradshaw, M. F. (1993) Vertical disparities, differential perspective and binocular stereopsis. Nature, 361, 253-255, 21 January.

Tittle, J. S., Todd, J. T., Perotti, V. J., and Norman, J. F. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. J. of Experimental Psychology: Human Perception and Performance, 21, 3, 663-678.

Wallach, and Karsh, E. (1963). The modification of stereoscopic depth-perception and the kinetic depth effect. American Journal of Psychology, 76, 429-435.

Wallach, H., Moor, M. E., and Davidson, L. (1963). The modification of stereoscopic depth-perception. American Journal of Psychology, 76, 191-204.

Woodson, W. E., Tillman, B. and Tillman, P. (1992). Human Factors Design Book. McGraw-Hill : New York.

Appendix A

In deriving the ISL, a distinction is made between viewing a scene comprised of real objects and stereograms that emulate 3-D objects. The considerations are somewhat different for the two cases though the underlying geometry is fundamentally the same (Cormack and Fox 1985), and is based on triangulation. For two real points with differential depth R, and one of which is fixated, the ISL had the following equation (Cormack 1984):

(1) $$R = (.5I(TAN(ATN(D/.5I) + .5d))) - D,$$

where I is the IPD, d is the retinal disparity in radians, and D is the viewing distance. For crossed disparity, d and R are negative. When the depth is small relative to the fixation, Equation (1) can be approximated by the equation (Graham 1965):

(2) $$R = \frac{D^2 * d}{I} .$$

Note that when the disparity is solved in this equation, it is scaled by the inverse of the viewing distance. This is where the ISL had received its name from. In the case of fusing the two half images of a stereogram, the background is usually taken as the reference plane (or the viewing distance). One point on the reference plane is fixated and determines the convergence angle. The simulated disparity is taken as the linear separation of a subfigure between the two half images. Thus, the ISL is given by the following equation (Cormack and Fox 1985):

(3) $$R = \frac{D * S}{I - S} ,$$

where S is the separation in cm and is negative for crossed disparity. In this context, the disparity in radian is S/D. Note that the same separation with different signs give different predictions of depth interval. Equation (3) was utilized for the theoretical predictions for the data obtained in this work.

APPENDIX A:
THE SPIN THEORY- A NAVIGATIONAL APPROACH TO SPACE PERCEPTION

1. INTRODUCTION

Visual space perception is a crucial feature for the survival of the organism. The theoretical questions that are addressed here are: How are the 2-D dynamic retinal images transformed into stable 3-D percepts of the object? Is our representation given in veridical or relative terms? and How can the objects' constancies be preserved? While theoreticians differ on these basic questions, the reality is that no integrating theory is widely accepted. This paper is mainly theoretical and elaborates on the Space Perception In Navigation (SPIN) theory, a navigational approach to space perception. The theory accounts for veridical depth perception and the preservation of objects' constancies (1-3), by combining visual with ocular and vestibular signals to create a stable representation across eye and head movements. The theory claims that *during fixations* the retinal signals are sufficient to reconstruct the egodistance and motion parameters of the eyes in space (1, 3), which enables object-centered representation (or passive navigation). Moreover, it also takes into account the Listing and Donders' laws, and the Vestibulo-Ocular-Reflex (VOR) to aid the navigational computations, and complement them when retinal signals are absent (active navigation).

2. VISUAL STABILITY AS TREATED BY THREE MAIN STREAMS IN PERCEPTION AND AUTOKINESIS

The visual world as described by Gibson (4-6) has the property of being stable and unbounded. This experience, according to Gibson, *is what a theory of perception must explain* (5). Attempts to address the perceptual stability can be classified into three main categories:

a) traditional inferential

b) Gibsonian direct perception

c) computational

The Inferential view regards visual stability as a large problem that is solved by several subsystems and with the aid of a great number of cues (see, e.g. 7). It states that we may perceive static environment as static depending on the outcome of the following basic mechanisms: 1. the elimination mechanism; 2. the translation mechanism (the take-into-account explanation); 3. the evaluation mechanism; 4. the calibration mechanism. For details about the explanations leading to visual stability, the reader may consult Gregory (8), Wertheim (9) or Bridgeman et al. (10). A major drawback in all inferential approaches is that they concentrate on egocentric representation and neither considers image instability due to head movements nor the differential retinal slips due to distance of objects from the eye. The direct perception approach, on the other hand, does consider these issues but does not provide a satisfactory formal account.

Wertheim (9) diagnosed many of these theoretical difficulties. He has advanced a model that assumes the existence of a *reference signal* which gets inputs from visual, ocular, and vestibular outputs. In my view, Wertheim considered all the relevant components for human navigation. The main problem with his model is that it is physically unsubstantiated. This is because visual stability, egomotion, and object's motion cannot be accounted for without the consideration of egodistance and structure of visual objects. These elements are missing in Wertheim's model.

In computational vision the problem of navigation of dynamic observer is analyzed by considering egodistance and the 6 motion parameters of the eye in space. In theory, absolute ego distances can be determined from the image data if the observer's motion parameters are known (1, 11). The question is whether the motion parameters and egodistances can be uniquely extracted from the retinal images without the aid of extraretinal signals. The consensus is that the answer to this question is negative (12). The most rigorous solution was advanced by Tsai and Huang (13) who show that 7 points and two views are required to recover the distance and the motion parameters (up to a scalar in the translation vector). Thus, all models in computational vision yield *relative depth perception* because their solution is up to a scalar. The problem is termed as the

indeterminate scale problem, and has direct implications to visual perception in general, and to Wertheim's model in particular.

To pinpoint our critique of models of the computational approach, we ask: what kind of perceptual experience would be predicted by the models when the input of the system is below the minimal conditions required for a solution, e.g. when the input to the system is a single point which creates the autokinesis phenomenon. The answer is that no solution will be obtained and the system will fail to stabilize the point. From a perceptual standpoint, the appearance of the point would be jittery, reflecting the unsteadiness of the eye.

Important observations related to autokinesis are that the effect can be obtained for more then one luminous point (14). In the case of two separated points, the torsional fluctuations are small even in the presence of large positional errors. Also, only small fluctuations in their angular separation are observed, which means that their rigidity is fairly preserved.

Thus the autokinetic phenomenon shows that the capacity of the human visual system is far beyond what was suggested by the most rigorous computational models. These models do not consider extraretinal signals. However, when extraretinal signals are considered, then on one hand, the indeterminate scale problem does not hold in normal vision because these signals may provide the missing information to the computational scheme. On the other hand, they cannot fully explain visual stability because the noise in these signals generates permanent slips of the retinal image which should be perceivable (15). Furthermore, retinal slips due to head movements can only partially be compensated because the vestibular system is insensitive to linear motion (see e.g. 16). Thus, one must conclude that a considerable amount of image instability should normally exist had visual stability been carried out on the basis of extraretinal information.

3. THE SPIN THEORY

The above drawbacks are settled within the SPIN theory as it regards space perception as a navigational process in which the organism continuously builds a stable representation of his

environment. This representation is given in terms of object's (exocentric) coordinates rather than retinal or head (egocentric) coordinates.

The fundamentals of the exocentric representation are illustrated in fig. 1a. The figure depicts the systems of coordinates that take part in the navigational computations. One is time-invariant system which is attached to the physical objects, with coordinates (X,Y,Z). The others include three dynamic systems of coordinates attached to the observer. One is head system and centered at the midpoint between the two vestibular organs, with coordinates (x_h, y_h, z_h). The other two are attached to the left and right eyes respectively, and centered at their corresponding *geometrical centers*, with coordinates (x_r, y_r, z_r; x_l, y_l, z_l) respectively. The equal distances between the origin of the head system and the corresponding centers of the two eyes (or the vector **f**) are considered here as having an effectively fixed length and assumed to be known to the processing mechanism. So are the radii of the eyeballs. In the model, these magnitudes are taken as a metric units, that serve to scale the representation. The eyes are approximated as spheres whose centers coincide with the corresponding eyes' centers as shown in fig. 1b. The optical apparatus of the eyes is approximated as pinholes which coincide with the corresponding reduced nodal points. The later points are taken as the perspective centers of the eyes.

Now consider that the following relations exist:

coordinate system = frame of reference

object's point = event

coordinates = representation

Then, linear algebra tells us that if the eye system has 3-D representation of objects' point and if the relative position between the eye and world coordinate systems is known, then the observer can calculate a representation of objects in the object frame of reference. This statement is equivalent to saying that the human perceiver can, in principle, builds in his "mind's eye" a 3-D image, or phantom, of the objects, and that the reconstructed image (in normal uncorrected vision) usually coincides (in numerical values) with the real location in space of the objects. This is the

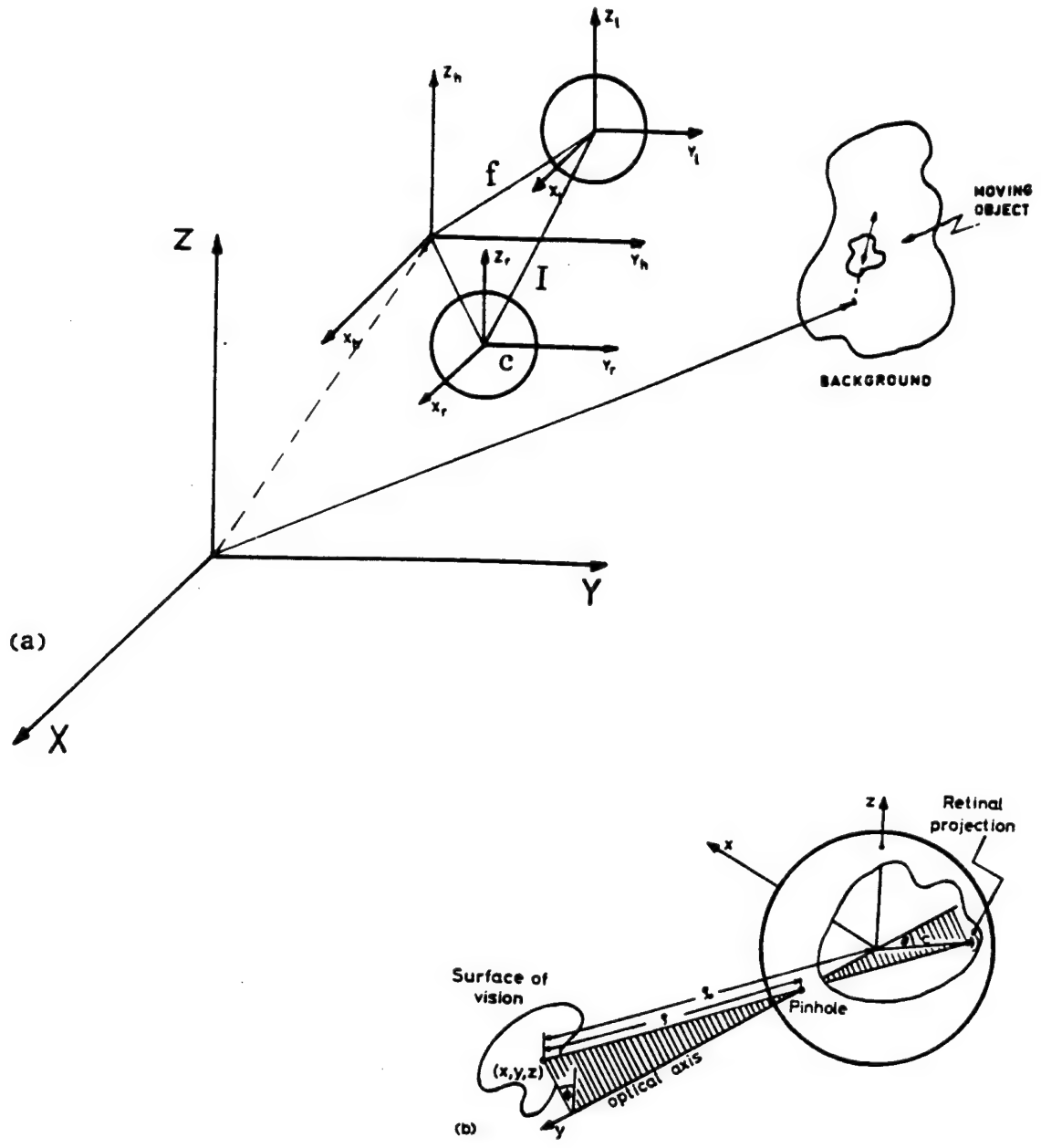


Figure 1a, 1b

The system of coordinates in the SPIN theory.

a: (X;Y;Z) is a system attached to the objects; $(x_h;y_h;z_h)$ is the head system; $(x_r;y_r;z_r)$ and $(x_l;y_l;z_l)$ are the system of coordinates attached to the eyeballs. I is the Inter Ocular Distance and **f** is the vector connecting between the head system and the eye system origins.

b: Definition of retinal coordinates of image points in the eye systems of coordinates - C is the radius of the eyeball. (From Hadani I, 1991.)

crux of the stability problem and the question is how the observer can attain egodistances and the knowledge of his relative position to object's world. As indicated earlier, this knowledge is derived either from the retinal information, or from extraretinal information.

A flow chart of the theory is given in fig. 2. The passive and active navigation parts are separated by a dashed line. The left part of the figure shows stages of the project-onto-the-retina and project-out-in-space of a stationary visual point. These involve the transformation of the point's spatial coordinates into retinal coordinates, followed by a stage of retinal motion extraction. The motion field is then processed to calculate the movement parameters of the eye in space and the egodistances. The motion parameters of the eye are then processed to calculate the project-out transform which is the inverse of the calculated project-in transform. The distance coordinate of the point with its retinal coordinates, accomplishes a 3-D retinal coordinate representation of the point. The 3-D retinal representation is projected-out-in-space at the output stage of the process. A detailed formal account of the project-in project-out stages and metric egodistance extraction, for an eye in pure rotations, is given in ref. 3.

The right part of the scheme in fig. 2 describes the stages which are involved in the processing of the extraretinal signals. The angular velocity of the eye relative to the head, taken from the oculomotor mechanism, are processed to give the angular position of the eye relative to the head. This stage is bound to the Donders' and Listing's laws. In parallel to the processing of eye-head movements, the head angular position in space is processed from vestibular signals. These serve to calculate the position of the eye in space. The linear displacements of the eye is calculated by double integration of the head linear acceleration, added with additional components caused by head rotations, because the eye and head system's origins are separated by the vector **f**. This stage of processing also has to consider the VOR. On the basis of our kinematical analysis, it is assumed that laws of eye movement and the VOR can assist the passive navigation and to veridical depth perception as will be suggested below.

The retinal and extraretinal signals are combined in the chart with two hypothetical analog xors units to give the value of the inverse transformation matrix at project-out-in-space stage. At present it is not clear exactly what are the rules by which these different sources of information are combined.

Vision research has revealed that the eye is basically an instrument for analyzing changes in light flux over time. Roughly speaking, without continuous changes in the light flux that is striking the receptors, there would be no neural response and the retinal image will disappear within 2-3 seconds (17). In accordance with this "limitation," human eyes are normally in continuous motion due to a variety of eye movements in their orbits, and due to head movements in space. These movements cause the retinal image to sweep and jump across the receptors, and to generate the desired local temporal changes of luminosity. Yet these movements are not perceived by the observer, and stationary scenes are perceived as such despite their retinal shifts. Furthermore, the magnitude of retinal displacement of an individual image point has a component which is determined by its distance from the eye, and is known as motion parallax.

The SPIN theory shows how the processing mechanism is assumed to handle this complicated situation for extended time periods. At $t=t_o$ the head system of coordinates coincides with the fixed-in-space system and the two eye-systems are in a primary position. Each point of the visible physical object is projected-onto-the-retina and gets a two-dimensional retinal label (coordinates) given in terms of the eye's system. Now suppose that at $t=t_1$ and (20 msec later, say) the eye had a small rotation in the orbit and at the same time the head had a small displacement. As a result of the compound movement, the point is displaced on the retina and gets a new local label. Knowing the angular velocity of that point, the following unknowns are estimated: a) the distance of that point, b) the parameters describing the momentary orientation of the eye-space systems, and c) the parameters describing the eye system translation which is also the head system translation added with the cross product of the head rotation and the vector **f**

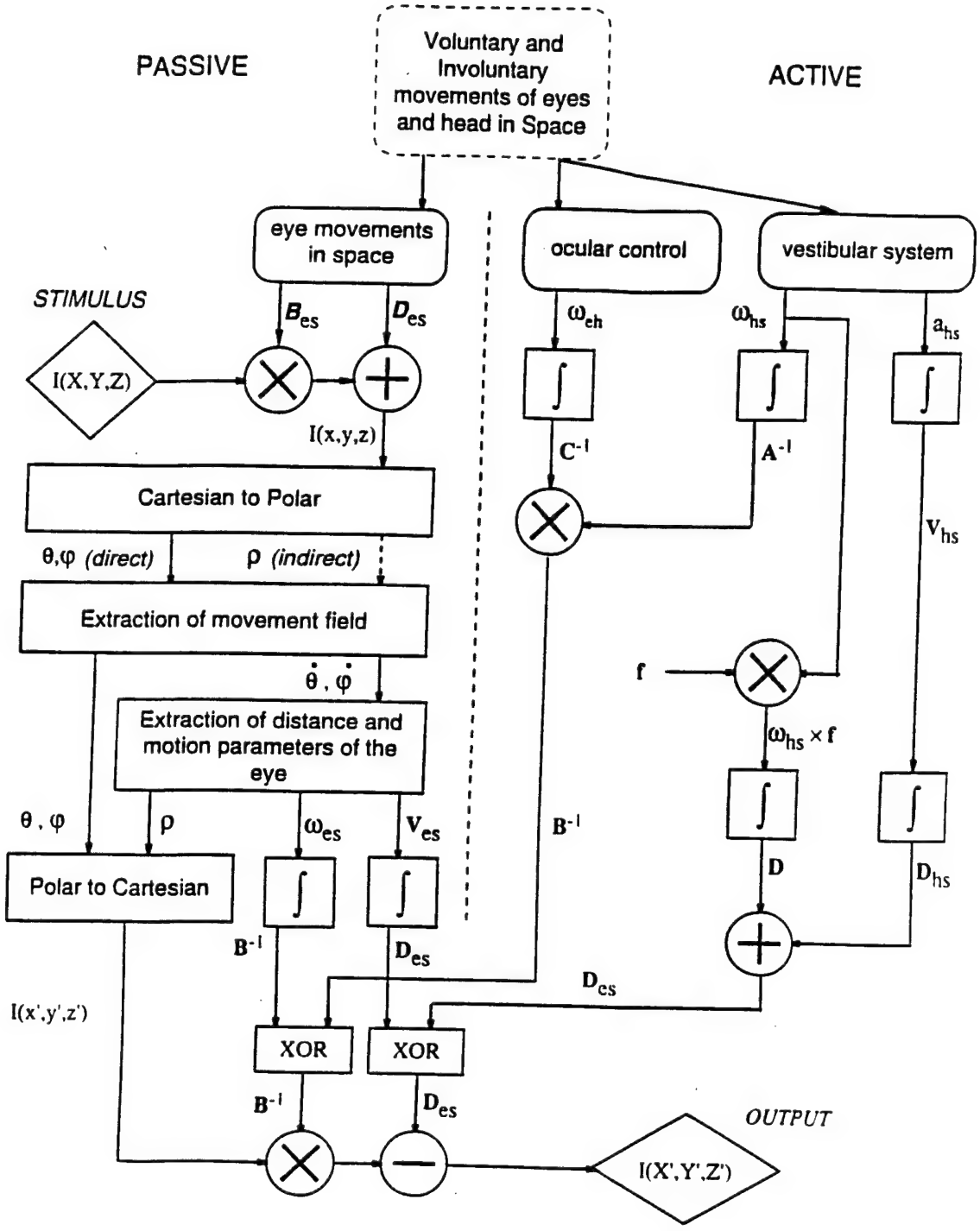


Fig. 2. A flow chart of the SPIN theory. For details see text.

(information about head rotation is taken from the vestibular system). The 3-D time-variant eye's coordinates (retinal location and estimated distance) are now transformed into 3-D **time-invariant** space coordinates by applying the inverse transformation. Integrating in time the estimated motion parameters enables the assumed processing mechanism to project the points out in space to its presumed veridical position. The processing mechanism will restart, or switch to a new frame of reference, when the observer attends another object which serves as reference, or blinks, or drastically changes his environment.

In cases of large and rapid eye movements (saccade), as well as head accelerations where retinal information is inhibited or smeared, the processing mechanism is assumed to utilize extraretinal signals to update the process of time-invariant representation.

The more general case is when there are several objects in the field of view and some of them are moving, the processing mechanism then selects one object as the spatial frame of reference. If that object is static, then the computations of the object-observer relations are correct and so are the movement trajectories of the other objects. However, when the selected object of reference happens to be one of the moving objects, then the calculated movement parameters of the observer would be biased (self motion may be wrongly inferred), and the stationary objects would be taken as moving.

A computer simulation of the project-in project-out process as applied to pure eye rotations (fixed head) is given in ref. 3. This simulation also emulated the Listing and Donders' Laws of eye movements, and autokinesis, and enabled some interesting predictions. First, it showed the autokinetic effect cannot be accounted for by noise in the retinal signals because such noise created large fluctuations of the perceived egodistance (including negative values) and relatively small positional drift. This implies that the retinal signals have to be more accurate than extraretinal signals. Second, autokinesis was best emulated by filtered white noise which emulates one that penetrate directly from ocular signals and affects the project-out process (fig. 3a). That negative distance values can be perceived has been reported by Skavenski, Haddad and Steinman (18) but

only when the eyeball was equipped with a molded contact lens, and when the lens was pulled sideways. This observation is in line with our conclusion that extraretinal signals that accumulate large positional errors may be the main cause of autokinesis. Third, the meaning of Listing and Donders' laws is that the ocular control mechanism probably utilizes only two independent parameters of rotation (19) when torsion is a function of the other two parameters. Under this condition the simulation has shown that for two separate points, the rigidity and no torsion are in agreement with psychophysical observations. When three independent rotational parameters were used in the simulation, large torsional drift in the autokinesis was obtained (not shown). Thus the lack of observed large torsional drift in autokinesis is attributed to the laws of eye movement. Fourth, the flow equations (3) indicated that the *structure of retinal motion field* is dictated by these laws. This was confirmed by the simulation as shown in fig. 3b. The figure depicts a polarized excursion of retinal projection of any visual point generated by an eye that obeys the Listing and Donders' laws. The pole of this excursion is the primary position. Additional aspects of extraretinal signals and their implications to the SPIN theory are given next.

4. EXTRARETINAL SIGNALS

4.1 Number of Ocular Signals

The eye movements receive their commands from brain centers that have the signal about the position and motion of the eyes relative to the head. Classical eye movement measurements led to the suggestions that there are five, or more, independent systems of ocular control. These suggestions are supposed to lead to a considerable amount of complication in any scheme that tries to elucidate the combination of extraretinal with retinal signals. However, recently Steinman et al. (20) argued that actually only two subsystems were diagnosed in their studies. One subsystem is a fast saccadic system and the other is a slower, smooth pursuit. These are used to fixate and track, respectively, a central representation of objects located in three-dimensional space. It is argued that this suggestion simplifies the hypothesized combination of retinal and extraretinal signals as suggested by the SPIN theory.

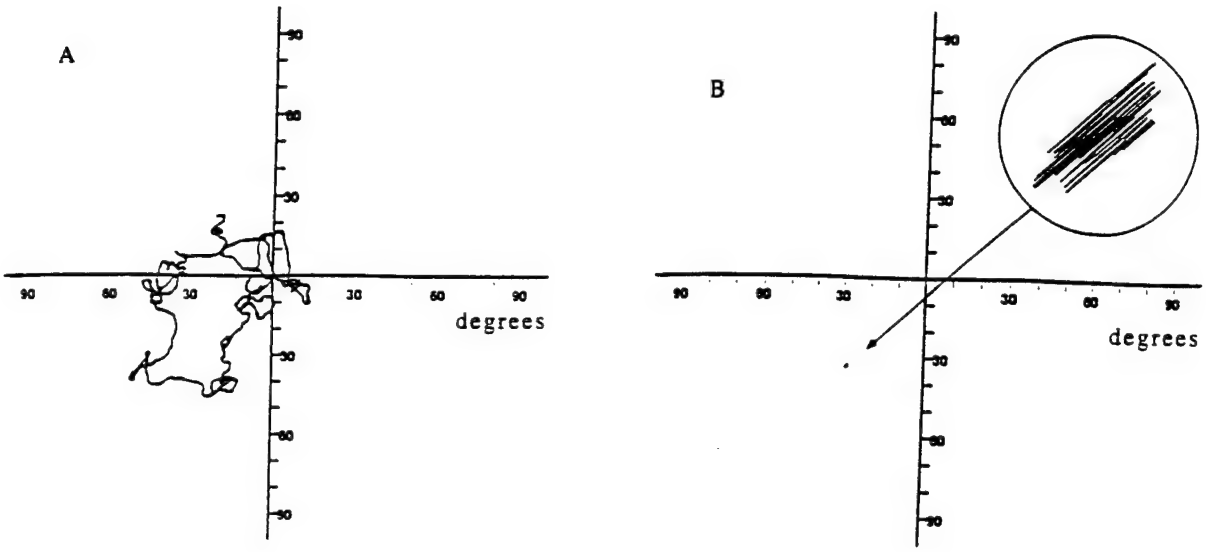


Fig. 3. (a) Autokinesis is simulated with 0.5% colored noise penetrating from extraretinal signals. The plot depicts the excursion of a point in 100,000 sampling units of 20 msec.

(b) Retinal excursion of a single point due to tremor of the eye. The magnified excursion of the point in the first 50 consecutive steps is shown in a circular window representing 16.8 arcmin in diameter. The polarization of the excursion steps is due to the implementation of the Listing and Donders' laws in the definition of the eye-head transformation matrix. The noise level is 0.01%. (From Hadani I, Ishai G, Frisch HL, Kononov A, 1994.)

4.2 The Laws of Eye Movements and Position Constancy

As mentioned above, motion trajectories of the eyes in their sockets are governed by the Listing and Donders' laws. The effect of the laws was used to establish the fact that the retinal image of any static object, for any given position of the head and fixation of the eye, falls upon identical retinal points. Thus, the laws became the basis for the constancy theory of localization (cf. 21), more specifically, to the oldstanding Lotze's theory of visual local sign. This theory assumed

that if the orientation of the head and the eye are known, then the position of a perceived object in relation to the vertical and horizontal of the visual field is determined by the retinal points simulated. Indeed, like all explanations in the inferential approach, what Lotze's theory is missing is the consideration of the head movements and egodistance. Then, if the observer fixates the same object twice when his head is translated between the first and second fixation, the object would not procure exactly the same shape and position on the retina.

There are increasing pieces of evidence that Donders' and Listing's laws are not due to orbital geometry or extraocular muscle mechanics. For example, they fail during sleep. Also VOR measurements in monkeys suggest that the laws prevail even during head movements, but with adjustments that are required for fixating a stationary point (22). Psychophysical measurements of the inclination of a vertical afterimage when the eye changes direction are in line, but deviate from those predicted by Listing's law (23). If these laws are not due to plant mechanics, then their existence raises the question: what kind of function do these laws fulfill? Some conjectures derived from our theoretical analysis relate them to the VOR as will be suggested below, but in general the answer to this question remains unclear.

4.3 Vestibular Signals

The vestibular organ does not inform us about the state of the external world (16), e.g. whether it is stationary or moving; rather it gives us information about the change of movement of our head relative to the inertial space. The response induced by the vestibular system typically occurs automatically, without any need of conscious control. Neurophysiological data have shown that interactions of inputs from different sensory system including visual, occur in neurons lying only one synapse from the central vestibular end organ (see e.g. 24). Also, the oculomotor output is intimately linked to the vestibular output. Thus, it is argued by Henn et al. (24) that if information from different sensory systems converge, the different inputs must have the same dimension for coding to be compatible. Similarly, compatible sensory coding is required for

appropriate motor responses that are based on multiple sensory inputs. As shown in sec. 2, the SPIN theory takes into account many of these aspects and combines them into a unified formal framework. One of the derivations of the theory is that if we visually select a stationary object as our frame of reference, then the calculated motion parameters of eye in space would be veridical and in line with those sensed by the vestibular ones because the latter are given in terms of the inertial space. However, when we select a moving object as our space (as in the case of vection) the calculations of our motion parameters will be correct with respect to the moving object but not with respect to the inertial space. Thus a conflict between retinal and extraretinal information is expected and presumably causes motion sickness.

4.4 Suggested Functions for VOR and Extraretinal Signals

Traditionally, the VOR is described as a distinct, phylogenetically old oculomotor subsystem, which serves to stabilize the gaze direction. It is supposed to act as a stereotyped reflex with definite input-output relations. On the basis of eye movement measurements in the dark, Collewijn (25) argued that the traditional view of the VOR is not realistic. He suggests a more fruitful hypothesis that the vestibular signals are just one of many inputs to a spatial localization process, which computes the relative position (and motion) between the subject and the target of his choice. In light of the later conjecture and Steinman et al. (20) suggestion about the existence of two subsystems of ocular control, and on the basis of our formal analysis of the passive navigation problem, I would like to suggest the functions of the VOR and extraretinal signals as follows:

a) The traditional function of stabilizing the gaze during head movements to allow fixations. This function may contribute to the navigational computations in several respects: 1) Fixating one of the visual points keeps the angle between the eye's momentary rotation and translation vectors perpendicular with respect to each other (which turns the velocity vector to lie in a plane perpendicular to the rotation vector. It reduces the number of degrees of freedom of the eye from 6 to 3 and simplifies the navigational computations). 2) Fixating one point also keeps the

retinal slips of other points small, and within range so that these slips may be captured by the (low level and local) retinal motion extraction mechanism. 3) In the general case of an eye moving in 6 independent degrees of freedom, the vestibular and ocular systems together may provide the computational mechanism information about the angle between the rotation and velocity vectors of the eye in space. Our analysis have shown that this is a minimal bit of information which is sufficient to solve the indeterminate scale problem.

b) To supply the navigational mechanism with the component of the head-motion-space in order to fill in the "dead" visual intervals created during saccades. This function enables the phenomenological continuity of vision across saccades and help to preserve the objects' constancies.

Crawford and Vilis (22) rotated monkeys sinusoidally at approximately .5 Hz and at amplitude of 60° while measuring their 3 rotational components of eye movements (the measurement device moved with the monkeys' head and the monkeys fixated a static target point). In agreement with function **a**1, they found that the pattern of eye position change was a necessary condition for stabilization of the gaze. Furthermore, they found that the quick phase of the VOR tilts out of the Listing's plane in a manner similar to saccades, which do so to obey the Listing's law, and with gains of .92, .87, and .66 for the horizontal, vertical and torsional axes respectively. Fixations with imperfect gains, particularly for torsion, may support conjecture **a**2 about orthogonality between the rotation and translation vectors even-though this condition is more constraining then the Listing and Donders' laws alone (the later refers only to rotations relative to the head). In addition, Crawford and Vilis (22) indicated that the position signals from the ocular motor-neurons perfectly matched to actual eye positions. This is a necessary input to the navigational mechanism assumed by the SPIN theory and is required to provide the initial conditions for the computation of egodistances after each saccade. Whether these conjectures, which were derived from pure theoretical considerations, hold empirically remain to be seen.

5. DISCUSSION

The SPIN theory regards space perception as a navigational process. Its essence can be extracted in following single key sentence: *The mind's eye observes a 3-D representation of the world given in world centered coordinates.* From this single sentence the rest of the arguments follow. First, it implies that the computations required for 3-D representation are carried out automatically and unconsciously. Therefore, our mind's eye does not observe the instability of the retinal image nor its distortion. Our mind's eye does not observe even the fact that the retinal image is inverted or smeared. Whenever we open our eyes, objects are perceived out-there-in-space. This representation is perceived only after the inverse transformation stage of the model (project-out) is carried out. Second, in order to obtain a stable representation, the theory considers eye movements in space which turn the traditional conception of the visual input, from a frozen picture to the optic flow. Third, projective geometry considerations indicate that the shift of any retinal point depends not only on the magnitude of eye movements but also on the distance from the eye to that point. Therefore, the theory considered the retinal motion, particularly the motion parallax, as the basic and reliable source of information for 3-D reconstruction. Here, we arrive at a basic theoretical difficulty because the problem of 3-D reconstruction was considered in the literature as indeterminate. This indeterminacy could be easily dismissed by assuming extraretinal signals as a reliable source of information about the position and motion of the eyes in space. But I have presented arguments which lead to the conclusion that instability of a static scene should be visible had the scene ultimately stabilized on the basis of extraretinal signals. Therefore, the main theoretical endeavor was invested in the mathematical derivations, and computer simulations to show that all the necessary information for 3-D reconstruction exists may in the optic flow as suggested by Gibson's direct approach (4-6). This effort turned out to be successful because we are able to show that the problem of passive navigation has a metric solution when we fixate. This solution was produced without the need of a priori knowledge about the objects, except for the assumption that they are static. Fourth, the theory assumes that in normal uncorrected vision a

static scene is perceived veridically. For that, the computation procedures needed at least one metric measurement unit. Three such candidates for metric units were identified by our analysis. In monocular vision these are: 1) the radius of the eyeball; 2) the distance between the eye-system and head-system origins. For binocular vision the metric unit is: 3) the inter-ocular distance. Fifth, empirical data and kinematical considerations indicate that vision is inhibited considerably during saccades. Therefore, the theory assumes that the extraretinal signals are utilized to update the navigational computations about the change of the eye position in space and enables the continuity in the visual experience. Moreover, the theory is mathematical and deterministic in nature and gives way to qualitative measurements in which calculations replace debate.

I consider the Wertheim and the computational models as competitive explanations to the SPIN theory. While I find Wertheim's model as a conceptually correct, it is interesting to note that its main drawback in this scheme is manifested by the computational models because the later are facing the indeterminate-scale problem. This problem means that one cannot decompose the retinal shifts into independent components of translation and egodistance. Thus, according to the computational models, one central element in Wertheim's model - the reference signal that encodes the eye motion in the inertial space - cannot be extracted uniquely from retinal information. In turn, the computational models yield only relative depth perception which does not preserve all of the object's constancies.

The SPIN theory remedies these drawbacks by postulating absolute depth perception. But the question is, what kind of empirical evidence lends support to this theory?

Even though absolute depth perception may not be reflected in human's conscious reports given in controlled laboratory experiments, there are many examples of behavior that indicate scaled depth perception. To cite a few representative cases, the ability of a rat to modify the force of his jump as a function of the distance between one stand and another (26) implies that the rat has metric perception of distance. As for humans, without scaled depth perception how can skilled athletes and other people precisely control the force and timing of their actions toward distant

targets? For example, a batter has to ensure that the bat meets the ball at a particular time and place.

Recent measurements on the ocular occlusion phenomenon have shown that in human observers this magnitude is about 11mm (27). Furthermore, when an eye in 6 degrees of freedom is considered, the distance between the origins of the eye and head systems presumably serves as an additional scale. Analogously, for binocular vision the distance between the eyes, known as the inter-ocular distance, is the scale. A psychophysical support for the latter conjecture was recently obtained by Hadani and Julesz by showing that the perceived depth in random dot stereogram was highly correlated with the observer's inter-ocular distance (28). Furthermore, depth judgments of 45 subjects fall nicely on the theoretical curve calculated from a formula that is based on triangulation. This finding has received additional support by neurophysiological measurements showing that cells in V1 of monkeys respond to the absolute depth of random dot stereogram (29).

Eye movement measurements with *free head* made by Collewijn et al. (30) have shown a high degree of accuracy of saccade, which is as good as the accuracy that is measured with fixed head, and there is a good eye-hand coordination with free head (31). These findings suggest the existence of a veridical scale in the visual pathways that enables accurate commands to the eye's (and other motor) muscles to reach their target. Accurate saccade with a free head also indicates that the oculomotor control mechanism takes into account the combined motion of the eye and head in space with the egodistance of the object. The significance of the Collewijn et al. (30) finding to the present arguments is manifested when one considers the fact that saccades are attention-driven *preprogrammed* movements of the eyes. As such, they are not guided by visual feedback and the accurate commands are essential for reaching (foveating) the target.

Given the above evidence, particularly the high correlation of stereoscopic depth judgment on the inter-ocular distance which suggests the latter as a scale in binocular vision, it is unlikely that the monocular mechanism utilizes an indeterminate-scale solution, as suggested by the common view. Thus, I suggest that the visual system as a whole reconstructs objects with a metric scale. The SPIN theory shows that this type of solution is mathematically feasible (3). Moreover,

the assumption made by the theory about the metric representation nicely solves the position and size constancy problem because if the mental representation is given in terms of object's coordinates undergoing linear transformation, then the position and the distance function are preserved.

ACKNOWLEDGEMENTS

The author wishes to thank Dr. Bela Julesz for his support, encouragement and inspiration, which made this work possible; and to Carol A. Esso for editing and competence in word processing. This project is sponsored by AFOSR Grant No. 93-NL-165.

REFERENCES

1. Hadani I, Ishai G, Gur M. Visual stability and space perception in monocular vision: mathematical model. Journal of the Optical Society of America 1980; 1:60-65.
2. Hadani I. The corneal lens goggles and visual space perception. Applied Optics 1991; 30:28:4136-4147.
3. Hadani I, Ishai G, Frisch HL, Kononov A. Two metric solutions to the three-dimensional reconstruction for an eye in pure rotations. Journal of the Optical Society of America 1994; 11:5:1564-1574.
4. Gibson JJ. The perception of the visual world. Boston: Houghton Mifflin, 1950.
5. Gibson JJ. The senses considered as perceptual systems. Boston: Houghton Mifflin, 1966.
6. Gibson JJ. The ecological approach to visual perception. Dallas: Houghton Mifflin, 1979.
7. Shebilske WL. Visuomotor coordination., visual direction and position constancies. In: Epstein W, ed. Stability and constancy in visual perception: Mechanisms and processes. New York: Wiley, 1977.
8. Gregory R L. Eye and Brain. London: Weidenfeld and Nicolson, 1966.
9. Wertheim AH. Motion perception during self-motion: The direct versus inferential controversy revisited. Behavioral and Brain Sciences 1994; 17:293-355.
10. Bridgeman B, van der Heijden AHC, Velichkovsky BM. A theory of visual stability across saccadic eye movements. Behavioral and Brain Sciences 1994; 17:247-292.
11. Negahdaripour S, Horn BKP. A direct method for locating the focus of expansion. Computer Vision, Graphics and Image Processing 1989; 46:303-326.
12. Koenderink JJ, van Doorn AJ. Affine structure from motion. Journal of the Optical Society of America 1991; 8:2:377-385.

13. Tsai RY, Huang TS. Uniqueness and estimation of 3-D motion parameters and surface structures of rigid objects. In Richards W, Ullman S, ed. Image Understanding 1985-86. Norwood NJ: Albex, 1985: chap. 6.

14. Hadani, I, Ishai G, Julesz, B. The Autokinetic movement and visual stability. Investigative Ophthalmology and Visual Science, 32(4), p. 900, Proc. of ARVO Meeting, March 1991.

15. Hadani I, Gur M, Meiri AZ, Fender DH. Hyperacuity in the detection of absolute and differential displacements of random-dot patterns. Vision Research 1980; 20:947-951.

16. Howard IP. The vestibular system. In: Boff KR, Kaufman L, Thomas JP, ed. Handbook of perception and human performance Vol. 1 - Sensory processes and perception. New York: John Wiley & Sons, Inc., 1986: 11-1 to 11-30.

17. Riggs LA, Ratliff F, Cornsweet JC, Cornsweet TN. The disappearance of steadily fixated visual test objects. Journal of the Optical Society of America 1953; 53:495-501.

18. Skavenski AA, Haddad G, Steinman RM. The extraretinal signal for the visual perception of direction. Perception and Psychophysics 1972; 11:4:287-290.

19. Koenderink JJ, van Doorn AJ. Method of stabilizing the retinal image, Applied Optics 1974, 13:4:955-961.

20. Steinman RM, Kowler E, Collewijn H. New directions for oculomotor research. Vision Research 1990; 30:11:1845-1864.

21. Boring EG. Sensation and Perception in the history of experimental psychology. New York: Appleton, 1942.

22. Crawford JD, Vilis T. Axes of eye rotation and Listing's law during rotations of the head. Journal of Neurophysiology 1991; 65:3:407-423.

23. Nakayama K, Balliet R. Listing's Law, eye position sense and the perception of vertical. Vision Research 1977; 17:453-457.

24. Henn V, Cohen B, Young LR. Visual-vestibular interaction in motion perception and the generation of nystagmus. Neurosciences Research Program Bulletin, MIT Press 1980; 18:4.

25. Collewijn H. The vestibulo-ocular reflex: An outdated concept? In: Allum JHJ, Hulliger M, ed. Progress in brain research. Elsevier Science Publishers B. V., 1989: 80:Ch 17.

26. Russell JT. Depth discrimination in the rat. Journal of Genetic Psychology 1932; 40:136-161.

27. Bingham PB. Optical flow from eye movement with head immobilized: 'Ocular Occlusion' beyond the nose. Vision Research, 1993; 33:777-790.

28. Julesz B. Dialogues on perception. Cambridge: MIT Press, 1995.

29. Trotter Y, Celebrini S, Stricanne B, Thorpe S, Imbert M. Modulation of neural stereoscopic processing in primate area V1 by the viewing distance. Science 1992; 257:1279-1281.

30. Collewijn H, Steinman RM, Erkelens CJ, Pizlo Z, Kowler E, Van der Steen J. Binocular gaze control under free-head conditions. In: Shimazu H, Shinoda Y, ed. Vestibular and brain stem control of eye, head and body movements, Springer Verlag, Japan Scientific Society Press, 1991.

31. Kowler E, Pizlo Z, Zhu GL, Erkelens C, Steinman RM, Collewijn H. Coordination of head and eyes during the performance of natural (and unnatural) visual tasks. In: Berthoz A, Graf W, Vidal PP, ed. The Head-Neck Sensory Motor System. New York: Oxford University Press, 1991.

APPENDIX B:
PERCEPTUAL CONSTANCY AND THE MIND'S EYE LOOKING THROUGH A TELESCOPE

1. INTRODUCTION

The geometric constancies of position size and shape, during head and eye movements in space, are not fully understood, even if one accepts any of the advanced theories of visual stability that were recently published in two target papers in the same issue of *Behavioral and Brain Sciences* (Bridgeman, van der Heijden and Velichkovsky 1994; Wertheim 1994). These explanations compete with the original concept of efferent copy that nulls the movement signals on the retina (von Holst 1954), yet leave some unsolved puzzles. How is it possible that static objects closer or farther from a static surround, appear at a standstill instead of moving faster or slower than the static surround? Furthermore, how is it possible that correcting glasses, prisms, and other prosthetic optical devices that slightly magnify or diminish the retinal image, do not affect the stability of the visual world after brief adaptation? Here, we propose that in order to achieve stability for objects at different distances from the viewer's eye, the objects may be represented as if viewed via a *1:1 magnification* telescope at a) *optical infinity*, as if the image projected on the mind's eye were collimated (Julesz 1995), or b) at their veridical distance (Hadani et al. 1978; 1980a). The first suggestion diminishes the stability problem for linear shifts of the eye and provides a way to solve the rotational component. The second suggestion treats both components of eye movements computationally.

The simplest nonmagnifying collimator is a telescope that requires two lenses of equal focal length (and of course, more than two compound lenses can both collimate the input and correct for many aberrations of the first lens). We propose that one of the lenses of the mind's eye is the optics of the eye; the cascaded lens is provided by the neural computation mechanism and encompasses a) the retinal and extraretinal signals that encode the motion of the eye and the head in space, and b) a cortical zooming mechanism that operates at anatomical mapping of various brain

areas from the fovea to V1, V2, and higher visual centers. Thus, the mind's eye can be conceived as a fictitious observer in the form of an array of spatiotopically mapped receptive fields, viewing the world through a 1:1 magnification telescope.

Visual stability as treated by three main streams in perception

The visual world, as described by Gibson (1950; 1966; 1979) has the property of being stable and unbounded. By stability, Gibson meant that the perceived world does not seem to move when one turns his eyes or his head around. By unboundedness, he meant that the perceived world does not seem to have anything like a visible circular or oval window frame. Thus, the phenomenal world seems to stay put, to remain upright, and to surround us completely. This experience, according to Gibson, is what a theory of perception must explain (Gibson 1966). Another difficulty for any theory of perception is the fact that the organism cannot scrutinize his retinal image directly, but only after some major transformations have taken place by early and unconscious mechanisms of visual processing. Pure projective geometry considerations tell us that the two-dimensional retinal image is optically inverted, distorted (due to the spherical structure of the eye), and constantly moving and jumping across the receptors (due to the combined motion of the eyes in the head and the head motion in space). Yet, neither these expected image distortions, nor the retinal slips (or the resultant blur) are perceived. Moreover, objects are also perceived monocularly out there and at different distances from the eye.

Attempts to address the problem of objects' constancies (or the problem of visual stability) can be classified into three main categories:

a) The traditional inferential approach

b) The Gibsonian direct perception approach

c) The computational approach

The traditional inferential view regards visual stability as a large problem that is solved by several subsystems and with the aid of a great number of cues (see Shebilske 1977). Different explanations suggest that we perceive static environment as static, depending on the outcome of the

following mechanisms: a) the elimination mechanism that uses subtraction; b) the translation mechanism (the take-into-account explanation) that uses compensation; c) the evaluation mechanism; and more recently, d) the calibration mechanism. (For a detailed discussion on the differences between the various mechanisms, see Shebilske 1977; Bridgeman et al. 1994.)

Two general notes on the inferential approach: First, most explanations concentrate on eye movements (pure rotations) relative to the head, leading, at best, to egocentric representation that is not time invariant. This drawback was remedied partially by the direct approach (Gibson 1966; 1979) and has received an important role in Wertheim's (1994) model. Second, the role of distance of objects from the eye in visual stability is underestimated. Therefore, depth perception and visual stability were treated separately.

The second category includes the Gibsonian direct perception and Wertheim's theory. Gibson (1950) has pointed out that the retinal projections contain considerably more information on the visual world than has been assumed by the traditional inferential approach. The transformations that successive retinal projections of the same rigid object are undergoing during locomotion of the observer contains some "higher-order" variables, that are candidates for the extraction of signals that could specify eye movements. Thus, the Gibsonian approach assumes that the perception of ego and objects' motion is derived exclusively from retinal afferent information, and explanation of visual stability does not need the concept of extraretinal signal (Wertheim 1994). Although Gibson (1966) recognized that vestibular and somatosensory afferent may also contribute to the percept of egomotion, in a later paper (Gibson 1968), he takes these signals as having only a role of confirmation. A major point of difference between the direct and the inferential approaches is that the former assumes that in normal vision, perception is veridical.

Attempts to bridge the gaps between the first two approaches include the Dual Mode model and the Wertheim model. The Dual Mode theory (Mack 1978; Matin 1986) has developed from concepts originally formulated by Wallach (Wallach 1959) to explain the phenomenon of center-surround induced motion. According to the Dual Mode theory, there exist two modes of visual

perception: a) A direct mode, in which extraretinal signals play no role, and that yields veridical percepts, and b) An inferential mode, which makes use of extraretinal signals, and that may yield illusions of motion. The theory assumes that there are two kinds of cues that may generate a percept of motion: object-relative and subject-relative cues. The object-relative cues stem from motion of objects relative to each other. The subject-relative cues stem from object motion relative to the observer. Wertheim (1994) presents evidence that shows the logic of the Dual Mode theory is flawed because the empirical criterion for distinguishing between the two modes is questionable. In turn, Wertheim (1994) presents a more comprehensive theory of visual stability in which retinal and extraretinal signals are combined and create a reference signal that encodes the movement of the retinal surface in the inertial space. An important point in Wertheim's theory is the recognition of the fact that extraretinal signals are partial or inaccurate and therefore, cannot serve as a reliable source for visual stability (Wertheim 1994; see also Howard 1986). The later argument provides the retinal signals a crucial role in establishing object constancies.

Common to most explanations of visual stability is the consideration of the visual stimulus as a homogeneously moving pattern in the fronto-parallel plane. Differential motion of surface points, as well as a surface's egodistance, is not usually considered. In our view, this oversimplification is a serious drawback of both approaches and leads to qualitative models of restricted nature. The drawback can be remedied by the rigorous computational approach where the frame of reference of the reconstructed objects is explicitly specified (see Hadani and Julesz's commentary in Wertheim 1994). However, computational analysis of the stability problem led to a different theoretical deadlock, which is the *indeterminacy* of the obtained solutions. The only determinate solution in the computational approach was obtained by a navigational theory of space perception advanced by one of us (Hadani et al. 1978; 1980a; Hadani 1991; Hadani et al. 1994) and is now called the Space Perception In Navigation (SPIN) theory (Hadani 1995). In general, the computational approach has no metaphorical window to elucidate what 3-D reconstruction or what shape from motion computations, means with respect to the mind's eye. Therefore, in the

following sections, we discuss various computational solutions and suggest the type of metaphorical window that can represent them.

Computational, navigational models differ from the two earlier approaches by also considering, within the context of visual stability, the egodistance and the 3-D structure of objects. (Hadani et al. 1978; 1980a; Longuet-Higgins and Prazdny 1980; Bruss and Horn 1983). Because of the complexity of the stability problem, the computational models are usually restricted to static (rigid) objects and the perceptual constancy problem is reduced to the analysis of the ability to recover from the optic-flow the six motion parameters of the eye (the reference signal), the egodistance, and/or the structure of objects. Bruss and Horn (1983) call this capacity Passive Navigation, a notion that we take as equivalent to the direct perception concept, with the exception that rigorous mathematics have been added. Several approaches are used to address the issue: the *discrete approach* (Longuet-Higgins 1981; Hadani et al. 1980a; Meiri 1980; Nagel 1981; Tsai and Huang 1985), the *differential approach* (Koenderink and van Doorn 1976; Longuet-Higgins and Prazdny 1980), and the *least squares approach* (Prazdny 1981; Bruss and Horn 1983). Works in the discrete and differential approaches are also characterized by analyzing the minimum conditions under which an ideal observer can solve the passive navigation problem. These minimal conditions are given in terms of a number of points and views. The most rigorous solution was advanced by Tsai and Huang (1984). They show that seven points and two views are required to recover the distance and the motion parameters (but up to a scalar in the translation vector). Longuet-Higgins and Prazdny (1980) show that the structure of object and motion parameters can be recovered from a single point and its near neighborhood (again, up to a scalar in the translation vector). Thus, the solutions in the computational approach face a scale ambiguity problem.

In contrast to the common view in computational vision, and in line with the concept of direct perception, the SPIN theory claims that the passive navigation problem has a unique solution, and this claim is supported by showing an existence proof (Hadani et al. 1994; Kononov 1996). Furthermore, the SPIN theory suggests three intrinsic "hard-wired" measures: a) the radius

of the eyeball, b) the fixed distance between the head system origin and the eye system origin, and c) the Interocular Distance - as metric units to scale the objects in physical units. The general solution offered by the SPIN theory for both passive and active navigation (without or with extraretinal signals, respectively) was possible because of the clear distinction that was made between retino-centric, head-centric and object-centric representations (Hadani 1995). The analysis attaches to each system a separate coordinate system and utilizes linear transformations to exchange representations between the different frames of references. For example, to solve the isolated passive navigation problem, there is no need to consider the head system, but only the eye and space systems as frames of reference. Then, the 3-D coordinates of objects are projected onto the retina through a pinhole and get a 2-D retino-centric representation. Optic-flow analysis is then applied to calculate the egodistance of each image point and the six motion parameters of the eye in space. The egodistance of each point, with its 2-D retinal location, accomplishes a 3-D retinotopic representation. Knowing the relative orientation between the eye system and the object system enables the projection of the retino-centric representation into objectcentric representation. The object-centric representation reflects the out-thereness of the perceptual experience. Furthermore, because linear transformations are invariant to the position of static points as well as to the distance function, then position, size and shape constancies are preserved in mental representation. Because this solution relies only on retinal information, a question is raised as to how such a system can confirm the veridicality of its solution. The consideration of this question in the context of normal and prosthetic vision led us to the statement of the magnification and distance paradoxes to be described in Section 4. Let us first examine, however, the meaning of visual perception offered by the three approaches as reflected by their metaphorical windows.

2. THE DIFFERENT WINDOWS METAPHORS AND THE TELESCOPE

To explain to novice students the essence of visual perception, investigators in the inferential approach introduce the concept of a projection plane in front of the observer (Goldstein

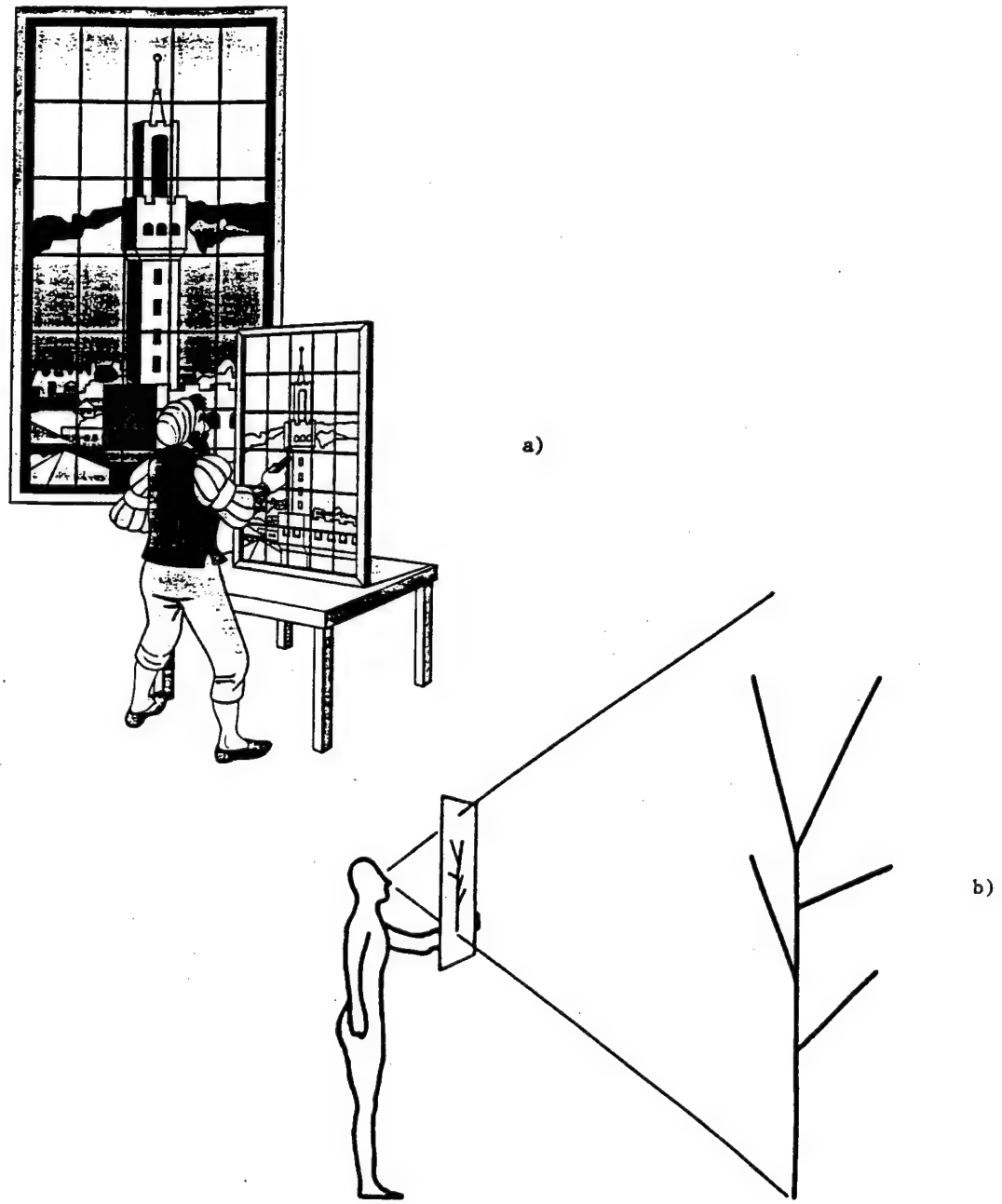


Figure 1. Two illustrations of Alberti's window metaphor.
(a) An artist painting a perspective using the Alberti's window method (From Goldstein 1996.)
(b) Schematic illustration of the Alberti's window utilized in a "structure from motion" treatise (From Cutting 1986.)

1996; Cutting 1986) as shown in Figures 1a and 1b. In Figure 1a, the window is static with respect to both space and the observer's vantage point. In our view, this ancient metaphor, called *Alberti's window*, is misleading. First, it does not account for many of the static features of the retinal image like spherical distortion, inversion, etc. Second, if the window or the observer moves, the position of the objects' representations in the window will change. Thus, position, size, and shape constancies are not preserved. Third, the image would be blurred by motion. Fourth, no veridical space perception can be recovered from a static image. Furthermore, this kind of interpretation of space perception leads to the most ambiguous theoretical situation where, in principle, an infinite family of spatial objects of different shapes and of different orientations may have the same retinal representation, or conversely the same retinal image may evoke the percept of an infinite number of real objects having different shapes and orientations as illustrated in Figure 2.

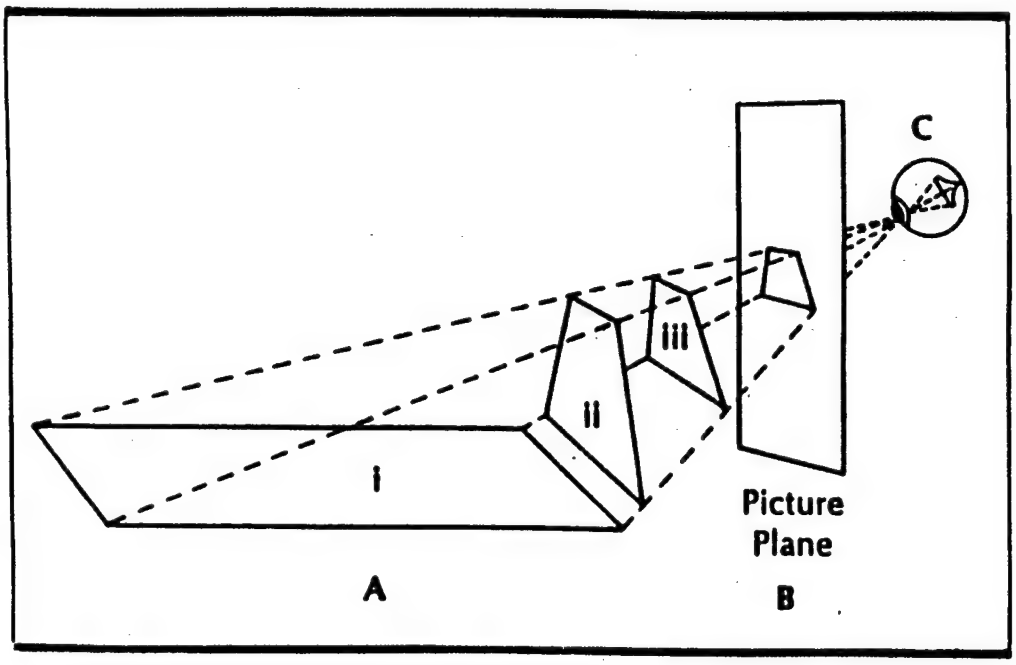


Figure 2. The common view in visual perception regarding the uncertainty in object reconstruction. Note: Different layouts (i, ii, iii) in different orientations project the same two-dimensional patterns in the picture plane B that represent Alberti's window, and consequently on the retina of the eye C. (From Hochberg 1971).

The analogous metaphor in direct perception attaches the window to the observer's head. Figure 3, which illustrates this metaphor, is depicted by Gibson as an upgrading of one made by Mach (1959), was entitled *The Visual Ego* (Figure 4). *The Visual Ego* metaphor better represents the dynamic nature of vision. Gibsons' improvisation of the Visual Ego is comprised of three snapshots (Figure 3). Each snapshot is made for a different heading direction. However, the three snapshots may not immediately reflect what Gibson meant by his conception of the mind's eye. Let us suggest our interpretation of these drawings. In our view, they demonstrate, in the first place, the "out-thereness" of the perceptual experience, e.g. that objects are perceived out there in space and not on some internal 2-D neural screen. Moreover, the reader may note that Gibson's nose is shown in all three drawings occupying the same position. Gibson's nose is an object that is attached to his head, and in this respect it is egocentered. The TV set shown in all three drawings is an external static object, but occupies different positions. Thus, the pertinent question here is: which of the two, the TV set or the nose, has a time invariant representation in Gibson's mind's eye? The reader can easily test and answer this question by rotating his head about a vertical axis while fixating, monocularly, a static object and attending to her nose. The reader may realize that the external static objects are normally perceived as static, while her nose is perceived as moving relative to the static object.

The meaning of the Visual Ego metaphor is that perception can be conceived as a process that turns the eyeball into a transparent window in front of the mind's eye, which is situated inside the skull. This analogy makes objects to be perceived as being stable and as to be viewed as out there and presumably, at their veridical distances. The transparent window metaphor apparently eliminates the need to compensate the retinal image shifts against head movements because the window is firmly attached to the head (though it still requires stabilization of the image against eye movement). This kind of representation, as the SPIN theory analysis can show, leads to egocentric (or head-centric) representation, which is time variant, and does not represent what we normally experience when we observe the world through the Visual Ego. As mentioned earlier, the visible

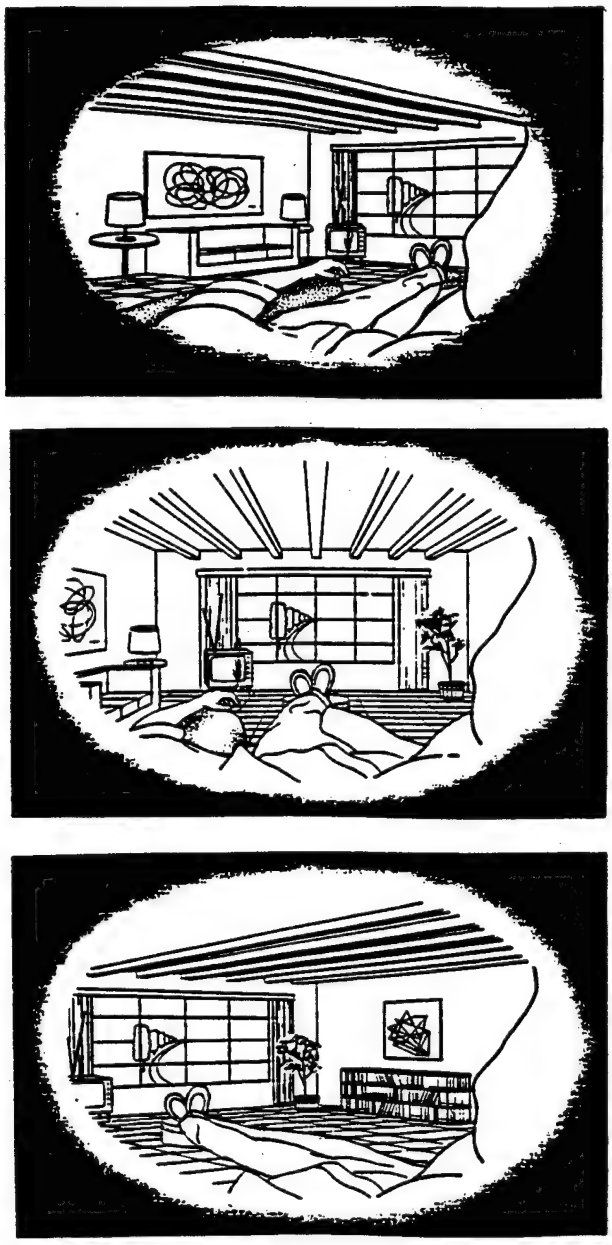


Figure 3. Gibson's improvisation of the Visual Ego metaphor originally advanced by Mach. The three drawings show three different snapshots of Gibson's study as seen by his mind's eye while sitting on a comfortable chair. Each snapshot is made for a different heading direction. The three snapshots demonstrate the "out-thereness" of the perceptual experience. Note that Gibson's nose is shown in all three drawings occupying the same position. The TV set, which is one of the external static objects, occupies different positions. The question is: Which of the two, the TV set or the nose, has time invariant representation in Gibson's mind's eye? The reader can easily test and answer this question himself by rotating her head about a vertical axis while fixating a static object and attending to her nose. The reader will realize that the external static objects are perceived as static while her nose is perceived as moving relative to the static object. (From Gibson 1979.)

Figure 4. The world seen by the left eye as depicted by Mach. This metaphor is named by Mach as the *Visual Ego*.

top of our nose is seen as moving relative to the stationary environment as would be expected in object-centric representation. Only with attentional effort can we perceive the objects moving relative to the nose, had the representation been head-centric. (Note that the ability to switch between the two modes expands the Dual Mode theory in the sense that different external objects - body centered, stationary, or moving - can be perceptually selected as our world's static frame-of-reference). In order to get a prolonged static (time invariant) percept of objects, one has to transform a retino-centric or head-centric representation that one may have at some stage of early computations, into object-centric representation. This can be done navigationally by integrating the "reference signal" (Hadani et al. 1994; Hadani 1995). Integration is obtained by updating the transformation matrices that describe the momentary relative position of the eye-head and head-space systems. Although Gibson was not aware of the available mathematical tools required to elucidate this point, we believe that he intuitively realized the objectcentric character of perception because he had indicated that direct perception is not characterized by egocentric representation (Gibson 1979, p. 201). Moreover, the SPIN theory disagrees with the interpretation given to direct perception by computational investigators regarding it as a look-up table (Ullman 1980). The SPIN theory analysis suggests that Gibson was theoretically correct about the notion of direct perception by showing that passive navigation can be obtained analytically. The question is whether it can be obtained empirically (see Discussion).

What are the analogous metaphors of the computational approach? Practically, the retinal image is taken as the effective window, as shown in Figure 5. However, the indeterminate solution means that on the sole basis of retinal signals we cannot perceive absolute depth, but only relative depth. Therefore, indeterminate solutions cannot achieve veridical object's constancies. The theoretical implication of this kind of indeterminacy is illustrated in Figure 6 by a family of parallel trapezoids. The reader should note that the degree of perceptual ambiguity obtained in the inferential approach (Alberti's window) is greatly reduced by the computational approach because the family of possible objects that can be reconstructed is limited to affine (or zooming)

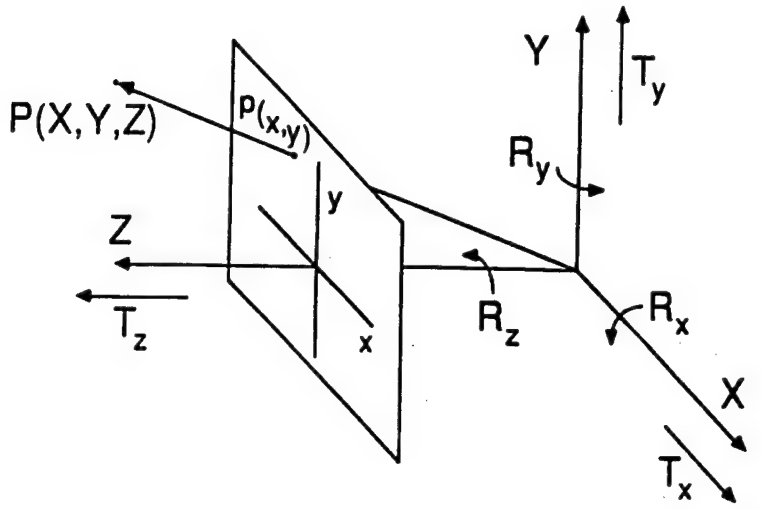


Figure 5. A standard window "metaphor" utilized by the computational approach to define the optic-flow. Note: The cartesian coordinate system is attached to the moving eye where origin is coinciding with the pinhole (exit pupil). Thus, the window shown is also attached to the eye and is positioned in front of the pupil. (From Simpson 1993.)

transformation. In other words, this means that we cannot perceptually distinguish between a near, but small object and far, but large object (Ullman 1979, p. 199; Bruss and Horn 1983). Moreover, it means that Wertheim's reference signal cannot be uniquely recovered and that visual stability can be obtained only with extraretinal information because they also encode the motion parameters of the eye in space. However, as noted earlier, the extraretinal signals are partial and inaccurate; thus, they cannot conceivably serve as a reliable source that supplement visual stability during fixations. They presumably serve the role of filling in the reference signal during saccades, e.g. when retinal signals are ineffective (see Hadani et al. 1994; Hadani 1995).

This discussion naturally raises the following question: Given that the retinal image is not a good metaphor for mental representation, what kind of metaphor could better represent the computational approach? As a first approximation, we would like to adopt the Visual Ego, suggested by Mach and elaborated by Gibson, for head movements. However, analysis of space perception in the context of prosthetic vision (Section 4), brought us to arrive at the telescope metaphor to be presented next.

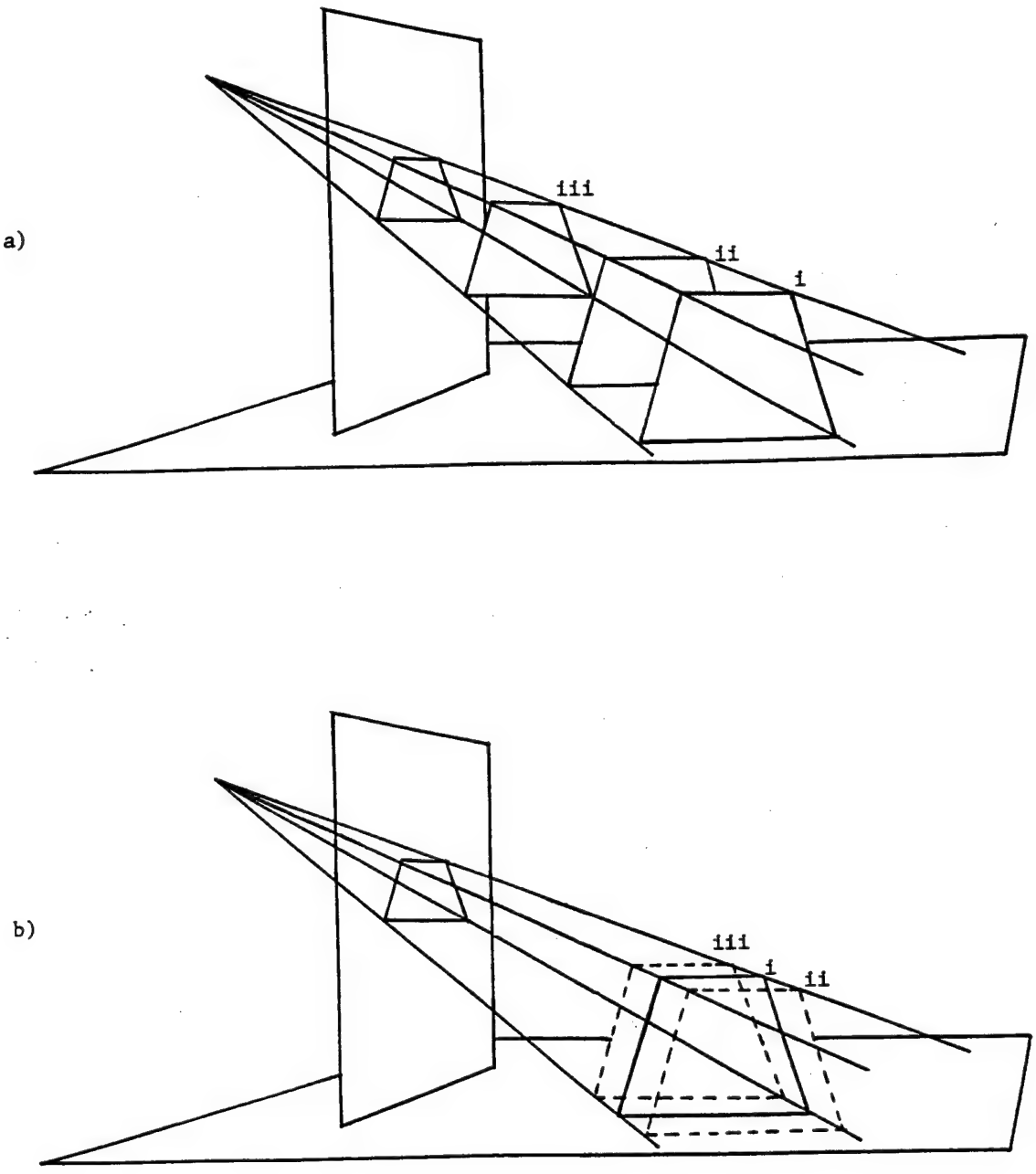


Figure 6. (a) The solid trapezoids (i, ii, iii) demonstrate the reduced uncertainty in object reconstruction obtained by the computational approach. Note: The three trapezoids represent an infinite family of parallel objects having a size scaled by the distance from the vantage point (compare with figure 2). (b) The degree of uncertainty, further diminished by the SPIN theory from infinite family of objects to a single object at the veridical depth (trapezoid i). Note: The ambiguity interval in reconstructing egodistance on sole basis of motion parallax cue is depicted by the dashed trapezoids (ii, iii) that are flanking trapezoid i.

3. THE TELESCOPE METAPHOR

The simplest nonmagnifying telescope requires two lenses of equal focal lengths (and of course, more than two complex compound lenses can both collimate the input and correct for many aberrations of the first lens). We propose that one of the lenses of the telescope is the optics of our eye; the cascaded lens is provided by the retinal and extraretinal signals and some neural circuits that can zoom (scale) the size of the perceived objects with their perceived distances. One of the features of these neural circuits may be the cortical magnification factor in anatomical mapping of various brain areas from the fovea to V1, V2, and higher visual centers that may provide several computational simplifications. The other feature is a neural zooming mechanism that operates at early stages of visual processing. The mind's eye, in this metaphor, is the collimated image at the output of the telescope that is captured by an array of spatiotopically mapped receptive fields.

The importance to space perception of the cortical magnification factor, in the form of a multiple logarithmic transformation, has been suggested by Schwartz (1980) in order to perform some affine transformations that seem to simplify operations that occur in form perception [e.g., retinal polar coordinates are transformed on V1 to the conceptually simpler cartesian coordinates]. Here, we argue that this factor, which may seem to introduce severe deformations in the representation of objects in the visual cortex, presumably, may play an additionally important role in visual stability by enabling a simple (similar) matching between two spherical systems that play a role in the control of eye movements: the fixed-with-respect-to-the-eye retinal polar system and fixed-with-respect-to-the-head orbital system. Both systems, being spherical, may be best represented with polar (or spherical) coordinates. Then, simple matching between the two systems can be obtained only when the eye is in the primary position and both systems coincide. However, when the eye maintains a secondary or tertiary position, there is no match between the two. Had the attention-driven, oculomotor-control mechanism utilized polar coordinates, more complicated computations to calculate the ballistic trajectories of saccades from one orbital position to another would be required. Much simpler computations may be required if the retinal and the orbital

coordinates are similar. This can be obtained, as suggested by Schwartz, by multiple logarithmic transformations converting both systems into cartesian. Then, rotations between the two systems are converted into translations.

The other feature, suggested as a mechanism for the cascaded lens is the neural zooming circuits. The conjecture that such mechanisms are used is supported by the size constancy phenomenon, stated as Emmert's law. Indeed, objects are perceived as having constant size even as their retinal image increases manyfold when they are viewed at reduced range. It can easily be shown with afterimages or with a tiny monitor attached to one's spectacles frame that the perceived image on the monitor, or the afterimage, shrinks as one inspects surfaces at increasingly closer ranges in order to achieve size constancy. Obviously, some cortical zooming mechanism at a rather early stage must perform this zooming operation. Furthermore, it has been suggested that infinitesimal affine transformations (also called Lie-germs, after the mathematician Sophus Lie) performed in the visual cortex can produce size and shape constancies (Hoffman 1970; Dodwell 1983) and emulated as a model on the Connection Machine by Papathomas and Julesz (1989).

Our suggestion that the retinal signals with the extraretinal signals and/or neural zooming circuits might be used to supplement the eye's optic by acting as a telescope in the viewer's perceptual system (what we call the mind's eye). We think this is novel and can explain many seeming paradoxes of visual stability. Whether this hypothesis can be experimentally confirmed remains to be seen, but as a possibility has some heuristic appeal.

4. THE MAGNIFICATION PARADOX

Analysis of realistic visual situations by utilizing the Visual Ego metaphor leads to several perceptual puzzles, particularly its inadequacy to account for prosthetic vision. These puzzles become more enigmatic when observers use low power prosthetic optical devices, such as correction glasses, that modify the effective magnification of the retinal image without affecting the stability of the perceptual experience. However, when the power of the optical device is high, such

as a binocular with 3x magnification or higher, the observer can hardly adapt to the impairment in stability (Demer et al. 1991). One of the aims of the present article is to state these puzzles and to suggest ways to resolve them with the help of the telescope metaphor. Because visual stability is basically a monocular phenomenon applicable to static objects, we will ignore problems of binocular vision and the perception of real motion.

Perceptual constancy, however, does hold under two puzzling circumstances:

A) When objects are in front or behind a reference plane (platform), and therefore, should move faster or slower than the platform, respectively, and with increased (or decreased) velocities as their distances from the platform become larger (or smaller), they are seen as stationary. For those workers in perception who do not believe in the percept of absolute depth, this perceived stability of the stationary, visual world - regardless of distance - is particularly enigmatic. We call this the *Distance Paradox.*

B) When optical devices are attached to the head in front of the viewer's eye with magnifications slightly larger or smaller than 1:1, one would assume that the stabilized platform at 1:1 magnification would drift faster (or slower) than the observer's head and eye movements as the optical magnification is larger (or smaller) than 1:1. Nevertheless, although new prescription glasses might cause momentary errors of visual stability, the visual system quickly adapts to the new glasses resulting in complete stability. Readers who wear bifocals, can verify that there is always a visible shear motion in the boundary of a bifocal lens. Only one side of the two half lenses can be stabilized at a time. We argue that perceptual phenomenon obtained in the unadapted aperture creates a *Magnification Paradox.*

In the sequel, we propose that both paradoxes can be resolved by the telescope metaphor a) when the retinal projection is a collimated and stabilized array; thus, effectively originating from a projection center at infinity, and b) when the objects are perceived at their veridical distances. We also propose that in normal uncorrected vision or after adaptation to an optical device, the effective magnification of the visual world at the mind's eye is 1:1 and the scene is perceived as stable in

spite of eye and head rotations. However, when the magnification of an optical device is high and cannot be adapted, the perceived distance of the scene changes too.

Some observations on visual stability with prosthetic vision

In order to better understand the magnification paradox, one may imagine a situation in which a visual scene is observed monocularly through a high magnifying monocle (1:3 or higher) that is attached firmly to the head (a monocle of a magnifying binocular can serve for that purpose). Alternatively, in the case of low power devices, where adaptation is fast, one may try to monocularly view the visual scene via a bipartite field (e.g. through both parts of bifocal glasses or through the border of simple correcting glasses). Then, rotating the head by swinging it left and right results in visible shear motion at the borderline. If the magnification imposed by the optical device is smaller than that of the unaided eye (as is the case with a single negative correction lens), then that part of the scene observed through the lens, and not adapted, will move in the same direction of the head movement while the other part will remain stationary. If the magnification imposed by the optical device is larger than that of the unaided eye (as is the case with positive correction lens), then the scene will swing in a direction opposite to the head rotation. When the whole scene is observed via a high power device, the magnified image cannot be stabilized by adaptation and the image will move in the opposite direction of the head. (The inverse effect occurs by inverting the binocular and observing the scene via the objective lens.) These image swings actually mean an impairment to the visual stability that any observer, with noncorrected vision, normally experiences. Furthermore, with bifocal or single-lens corrective glasses, one can easily observe that it is possible to stabilize either part of the visual field voluntarily and almost without practice, but it is impossible to stabilize both parts simultaneously.

It can be shown that an observer can adapt to different whole field magnifications introduced to the two eyes. This capacity indicates that the visual system can adapt to small

magnification variations, and that this adaptation is probably carried out by independent monocular mechanisms.

Having pointed out situations in which a magnifying device impairs visual stability in one direction of the head rotation and an image diminishing device impairs stability in the opposite direction, let us define *unity magnification* as the magnitude of the apparent magnification of the perceived scene viewed with a normal, unaided eye when no impairment of visual stability is observed. The last definition is different from the magnification of the optical apparatus of the eye - the cornea and the lens - in the sense that it refers to the apparent magnification of the mind's eye. Adopting the definition of unity magnification, one can visualize the mind as a fictitious observer (say, an array of spatiotopically mapped receptive fields) situated inside the skull, observing the world through a unity magnification telescope. Unity magnification means that the scene observed by the mind's eye is not distorted as would be expected from the high refracting power of the eye's optics and from the cortical magnification factor. Indeed, unity magnification also means that the angular velocity of the perceived image is opposite but equal to the angular velocity of the head. Then, for a stable scene, the ratio of the perceived image-angular velocity to the head-angular velocity equals one and the difference between the two (the amount of instability) equals zero.

Now, imagine a situation where the head of a human observer rotates by an integral rotation about a vertical axis while the eye maintains the primary position at the beginning and at the end of the head rotation. Then, pure kinematical considerations show that the observer's eye rotated exactly by an integral rotation in space about a vertical axis (for better realization of the physical situation of integral rotation of the eye in space, one may consider the observer's eye fixed to his or her head throughout the whole rotation). The magnification paradox becomes obvious when one considers two cases of a rotating monocular observer: a) the observer has normal uncorrected vision and is equipped with a cross-hair in front of her eye; the cross-hair is attached to her head and centered at the optical axis in the primary position, and b) the observer is equipped with a magnifying device having a reticle with a central cross-hair. In case a) the observer should

not experience any image instability. In case b), however, the visual scene observed, via the monocle, is perceived to be continuously moving relative to the head. As we indicated earlier, the magnitude of the image movement is sign dependent and is a monotonic function of the magnifying index of the monocle (Case b). It should be zero for the unaided eye (Case a). However, in both cases, when the integral head rotation is accomplished, he or she will see, coinciding with the cross-hair, exactly the same visual point that was overlapped by the cross-hair at the initiation of the rotation. The enigma here is that a spatial point viewed via a magnifying device with the head performing one integral rotation (that makes the point to be perceptually moving via the metaphorical window at an angular velocity different than the head), ends up with perceiving the same point at the same position, as when the head rotation had been started. This result is in contrast with the expectation that the positional error of the unstable image will reset itself as some point of the integral rotation, or alternatively accumulate and create a positional error at the end of the rotation. Yet, the reader can verify that no reset occurs and positional error is not integrated. Thus, image instability creates a magnification paradox.

5. DISCUSSION

Essential differences between three approaches in cognitive psychology, all dealing with the problem of perceptual constancy, were examined in this chapter with the help of their representative metaphors. This examination required, in the first place, identification of the representative metaphor that best represents each of these approaches, particularly the direct perception and the computational approaches. Thus, our choice may be disputable and needs justification. The Alberti's window did not create a problem because it is extensively used in old and modern treatises of sensation and perception (Goldstein 1996). As for the visual ego, we argue that although Gibson himself used Alberti's window in his latest book to illustrate the ambiguity that perspective projections creates to vision (Gibson 1979), we believe that the visual ego is a better representation of Gibson's view, at least it is at the latest evolutionary stage of his theory.

Moreover, a different improvised Mach's visual ego appeared in his 1950 and 1966 books, and indicates that this provocative metaphor of the mind's eye attracted his attention and motivated him to add his own interpretation that also appeared in his 1979 book (Figure 3). Our choice of the retinal image as a tentative metaphor for the computational approach was based on the standard definition of the optic flow (Bruss and Horn 1993; Simpson 1993). However, as indicated earlier, the nature of the effective window, after the computations are carried out, remains ambiguous, leaving room to suggest the telescope metaphor as one that may inspire the computational approach.

The telescope metaphor has all the advantages of the ego metaphor and more. Being attached to the head, it emulates the need to compensate for rotational head movements (although, it does not explicitly manifest the need to compensate for eye movements relative to the head). The input to the mind's eye in this metaphor is taken as the image that is created by the cascaded neural lens at the exit of the telescope. Thus static objects in front of a static platform are seen as static when the eye moves because motion parallax is diminished for this collimated image - and may provide an answer to the distance paradox. The zooming mechanism emulates adaptation to correcting glasses, prisms, and other optical devices, making the metaphor a suitable model for both normal and prosthetic vision. It may provide an answer to the magnification paradox with the assumption that the ratio of the perceived and veridical distances equals the magnification of the telescope, and that unity magnification may yield veridical perception of distance. Moreover, a standard telescope generates linear shifts, rotations, and zooming, that are analogous to global affine transformations of translation, rotations, and dilation/contraction, respectively. Thus, visual stability in this model can be obtained by compensating just for this subgroup of the Lie transformations. Indeed, all these features of the telescope are presumably accomplished by the hypothetical cascaded lens that emulates neural processing of retinal and extraretinal signals. In conclusion, the telescope metaphor suggests that the perceptual problem of visual stability (or object constancy) can be metaphorically reduced to *geometric optics*.

The consideration of the metaphorical windows, as a reference for comparison between the three approaches, provided a simple way to illustrate the differences in the uncertainty of object reconstruction embedded in the different models. This uncertainty is shown to be the greatest in the inferential approach, it is considerably narrowed by the computational approach, and diminished by the direct perception approach. In the SPIN theory, it is explicitly reduced to a measurement error. In this context, we note that Ullman's interpretation of the ambiguity entailed in the computational approach, e.g. that "we cannot perceptually distinguish between a near but small object and far but large object" is correct but may have additional meaning. Here, we would like to suggest an alternative theoretical interpretation to the scale ambiguity in the computational approach which is more relevant to the core issue of this chapter. The scale ambiguity in the system of equations, derived by the computational models, stems from the fact that the number of unknowns is greater (by one) than the number of equations. Yet, a navigational-computation scheme (integration) can be applied even with this handicapped system of equations by assigning, as an initial condition, an arbitrary value for one of the unknowns such as egodistance or motion. The interesting consequence of this "trick" is that the system would produce a solution that retains position size and shape constancies across time. However, the reconstructed world and the motion parameters given by this solution would not be veridical except when the initial value assigned is veridical. Moreover, such a system may retain this nonveridical solution without sensing its contradiction with the physical world. An active system, on the other hand, can use the vestibular signals that are anchored to the inertial space to confirm the veridicality of its solution (Gibson 1968) and to set the initial value to a veridical one. In humans, when a disparity between the visual and vestibular signals exists, they evoke the physiological unpleasant feelings of motion sickness. The fact that we normally do not experience motion sickness may suggest that in normal vision, there is an agreement between visual and vestibular signals and perception is veridical.

The previous arguments tap the crux of the difference between the views of the two authors, as were reflected in the two versions of the telescope metaphor. The first view (BJ) is in

agreement with the consensus in computational vision about the perceptual implications of the scale ambiguity problem. Thus, this view adheres to the perception of relative depth. The second view (IH) is grounded on the SPIN theory's derivations and agrees with Gibson's claim about the veridicality of space perception. Thus, this view adheres to the notion of absolute depth perception. Indeed, this kind of dispute should be resolved empirically by conducting a crucial experiment that will decide between the two views. Because no such experiment is yet envisioned, we would like to present an additional observation involving prosthetic vision that may lead to the design of such a crucial experiment. The suggested observation becomes meaningful with the following two premises: a) The same object cannot simultaneously occupy two different locations in space, and b) In prosthetic vision, the perceived distance of a scene varies with the magnification of the device. The observation requires a simple negative lens (a correcting glass for myopia may suffice) that is mounted on one's head a few centimeters in front of the eye and covers only part of the visual field. The scene should be static. Due to the optical interference, some of the objects viewed close to the border of the lens, create double images (e.g. monocular diplopia). The first question raised by this observation is: Which of the two perceived double representations of the same object better represents its veridical position? Applying small head rotations may reveal that one, or both, of the images may be perceived at different egodistances and as drifting contingent with the head rotations. As noted earlier, it is likely that one of the double images may be adapted to yield a stable percept. Thus, the second question is: What is the meaning of the stability perceived under these circumstances? Or equivalently, what is the perceptual meaning of the notion of unity magnification? The significance of this kind of observation in resolving the dispute of relative-absolute depth perception stems from the fact that the extraretinal signals for the diplopic object are identical.

In conclusion, Alberti's window is an ancient metaphor that does not faithfully represent that state of the art in perception. Gibson's version of Mach's inspiring visual ego metaphor better represents the dynamic nature of vision and the problem of perceptual constancy. The newly

derived and more sophisticated telescope metaphor may account for both normal and prosthetic vision and is suggested here as a challenge for the computational approach. Our observations on perception, using prosthetic vision, lead to certain paradoxes that should be resolved empirically and may provide an answer to the dispute on relative-absolute depth perception.

ACKNOWLEDGEMENTS

The authors wish to thank Alex Kononov for collaboration in the mathematical derivations of the SPIN theory; to Drs. Harry L. Frisch and Gideon Ishai for consultation on theoretical issues; and to Carol A. Esso for editing and for her competence in word processing. This project is sponsored by AFOSR Grant No. 93-NL-165.

REFERENCES

Bridgeman B, van der Heijden AHC and Velichkovsky BM. (1994). A theory of visual stability across saccadic eye movements. *Behavioral and Brain Sciences* 17, 247-292.

Bruss AR and Horn BKP. (1983). Passive navigation. *Computer Vision, Graphics, and Image Processing* 21, 3-20.

Cutting JE. (1986). *Perception with an Eye for Motion*. Cambridge: MIT Press.

Demer JL, Goldberg J, Franklin IP and Schmidt K. (1991). Validation of physiological predictors of successful telescopic spectacle use in low vision. *Investigative Ophthalmology & Visual Science* 32:10, 2826-2834.

Dodwell PC. (1983). The Lie transformation model of visual perception. *Perception and Psychophysics* 34, 1-16.

Gibson JJ. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.

Gibson JJ. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Gibson JJ. (1968). What gives rise to the perception of motion? *Psychological Review* 75, 335-346.

Gibson JJ. (1979). *The ecological approach to visual perception*. Dallas: Houghton Mifflin.

Goldstein EB. (1996). *Sensation and Perception*. Brooks/Cole: Pacific Grove.

Hadani I. (1991). The corneal lens goggles and visual space perception. *Applied Optics* 30:28, 4136-4147.

Hadani I. (1995). The SPIN theory-a navigational approach to space perception. *Journal of Vestibular Research*, 5, 6, 443-454.

Hadani I, Ishai G and Gur M. (1978). *Visual stability and space perception in monocular vision: A mathematical model*. Technical Report TSI 07-78, Technion-IIT, The Julius Silver Institute of Biomedical Engineering Sciences, Haifa, Israel.

Hadani I, Ishai G and Gur M. (1980a). Visual stability and space perception in monocular vision: Mathematical model. *Journal of the Optical Society of America* 70:1, 60-65.

Hadani I, Gur M, Meiri AZ and Fender DH. (1980b). Hyperacuity in the detection of absolute and differential displacements of random-dot patterns. *Vision Research* 20, 947-951.

Hadani I, Ishai G, Frisch HL and Kononov A. (1994). Two metric solutions to the three-dimensional reconstruction for an eye in pure rotations. *Journal of the Optical Society of America* 11:5, 1564-1574.

Hochberg J. (1971). Perception II. Space and Movement. In: Kling JW and Riggs LA (Eds.). *Woodworth and Schlosberg Experimental Psychology* 3rd edition, 475-550, New York: Holt.

Hoffman WC. (1970). Higher visual perception as prolongations of the basic Lie transformation group. *Mathematical Biosciences* 6, 437-471.

Howard IP. (1986). The vestibular system. In: Boff KR, Kaufman L, and Thomas JP (Eds.). *Handbook of perception and human performance Vol. 1 - Sensory processes and perception*. New York: John Wiley & Sons, Inc., 11-1 to 11-30.

Julesz B. (1995). *Dialogues on Perception*. Cambridge: MIT Press.

Koenderink JJ and van Doorn AJ. (1976). Local structure of movement parallax of the plane. *Journal of the Optical Society of America* 66:7, 717-723.

Kononov A. (1996). *SPIN theory and indeterminate scale problem*. Unpublished Master's thesis, Rutgers University, New Brunswick, NJ.

Longuet-Higgins HC. (1981). A computer algorithm for reconstruction a scene from two projections. *Nature* 293, 133-135.

Longuet-Higgins HC and Prazdny K. (1980). The interpretation of moving retinal images. *Proceedings of the Royal Society of London B* 208, 385-387.

Mach E. (1959). *The analysis of sensations*. New York: Dover Publications, Inc.

Mack A. (1978). Three modes of visual perception. In: Pick HL and Saltzman E (Eds.). *Models of perceiving and processing information*, 171-186, Hillsdale, NJ: Lawrence Erlbaum Associates.

Matin L. (1986). Visual localization and eye movements. In: Boff KR, Kaufman L and Thomas JP (Eds.). *Handbook of Perception and Human Performance. Vol. I: Sensory processes and perception*, ch. 20, New York: Wiley.

Meiri Z. (1980). On monocular perception of 3-D moving objects. *Institute of Electrical and Electronics Engineering Transactions on Pattern Analysis and Machine Intelligence*, PAMI-2, 582-583.

Nagel HH. (1981). On the derivation of 3D rigid point configurations from image sequences. Paper presented at the IEEE Conference on Pattern Recognition and Image Processing, Dallas, Texas, August 1981.

Papathomas TV and Julesz B. (1989). Lie differential operators in animal and machine vision. In: Simon JC (Ed.). *From the pixels to the features: Proceedings of the COST 13 Conference*. North-Holland, invited chapter, 115-126.

Prazdny K. (1981). Determining the instantaneous direction of motion from optical flow generated by a curvilinear moving observer. *Computer Graphics and Image Processing* 17, 238-248.

Schwartz EL. (1980). Computational anatomy and functional architecture of the striate cortex: A spatial mapping approach to perceptual coding. *Vision Research* 20, 645-669.

Shebilske WL. (1977). Visuomotor coordination., visual direction and position constancies. In: Epstein W (Ed.). *Stability and constancy in visual perception: Mechanisms and processes*, 23-69. New York: Wiley.

Simpson WA (1993) Optic flow and depth perception. *Spatial Vision* 7, 1, 35-75.

Tsai RY and Huang TS. (1984). Uniqueness and estimation of 3-D motion parameters and surface structures of rigid objects. In: Richards W and Ullman S (Eds.). *Image Understanding 1984*, 135-172. Norwood NJ: Albex.

Ullman S. (1979). *The interpretation of visual motion.* Cambridge: MIT Press.

Ullman S. (1980). Against direct perception. *Behavioral and Brain Sciences* 3, 373-415.

von Holst E. (1954). Relations between the central nervous system and the peripheral organs. *Animal Behavior* 2, 89-94.

Wallach H. (1959) The perception of motion. *Scientific American* 201, 56-60.

Wertheim AH. (1994). Motion perception during self-motion: The direct versus inferential controversy revisited. *Behavioral and Brain Sciences* 17, 293-355.

APPENDIX C:
SIMPLE INTEGRATIVE METHOD FOR PRESENTING HEAD CONTINGENT MOTION PARALLAX AND DISPARITY CUES ON INTEL PROCESSOR BASED PCS

Janos Szatmary, Itzhak Hadani, Bela Julesz

ABSTRACT

Rogers and Graham (1979) developed a system to show that head movement contingent motion parallax produces monocular depth perception in random dot pattern. Their display system was comprised of an XY oscilloscope driven by function generators, or a special graphics board that triggered the X and Y deflection of a raster scan signal. Replication of this system requires costly hardware that is no longer on the market. In this paper we show how the Rogers and Graham method can be reproduced with any Intel processor based PC with no hardware cost. An adapted joystick sampled by the standard game-port can serve as a provisional head movement sensor. Color monitor resolution in displaying motion was effectively enhanced 16 times by the use of anti-aliasing enabling the display of arrays of 1000 random dots in real time (double buffer) with a refresh rate of 60 Hz and above. The color monitor enabled the use of the anaglyph method, thus combining stereoscopic and monocular parallax manipulations on a single CRT display without refresh rate cost. The power of our display is demonstrated by a psychophysical measurement in which subjects nulled head-movement-contingent illusory parallax, evoked by a static stereogram, with real parallax. The amount of real parallax required to null the illusory stereoscopic parallax monotonically increased with disparity.

INTRODUCTION

Rogers and Graham (1979) were pioneers in advancing a dynamic display system showing that head movement contingent motion parallax produces a strong monocular depth percept in a random dot pattern. The unique characteristic of a random dot pattern is that this stimulus is devoid of monocular cues for depth as to the form or shape of the 3-D structure, by analogy with depth percept evoked by stereoscopic depth in Random Dot Stereogram (RDS) (Julesz 1960, 1971). This method indicates that motion parallax information is a sufficient condition for monocular depth perception. The original Rogers and Graham (1979) display system emulated motion parallax through the use of an XY oscilloscope driven by function generators that produced the X and Y deflection of a raster scan signal. Lateral head motion of the observer was sensed by a potentiometer yoked to a sliding chin rest. The latter signal was used to add a vertical waveform (sine, square or sawtooth) amplitude modulation to the X deflection of the dots. The modulation of dot position allowed the subject to observe horizontal monocular depth corrugations of the pattern in the fronto-parallel plane. Later replications of the system required, in general, costly hardware, or electronic-shop work. For example, a special graphics board to generate in real time a dynamic random dot pattern (Rogers and Rogers, 1992), an analog-to-digital interface card to sample the voltage on a potentiometer that senses the head movements (van Damme and van de Grind, 1993), or adaptation of the raster scan of a video monitor that enables appropriate modulation of the X deflection (Ono *et al.* 1986). These systems also restricted the shapes of motion parallax fields that can be presented. Only a few systems combine stereoscopic information with monocular motion parallax. For example, Rogers and Collett (1989) combined the fields of two black and white monitors with a haploscope. Similarly Ichikawa and Saida (1995) utilized polarizers and two black and white monitors whose fields were superimposed by beam spitters to add binocular disparity cues. The latter method involved off-line generation of 26 pairs of random dot stimuli stored in a frame memory. In this paper we show how the Rogers and Graham method can be reproduced with any Intel processor based PC with almost no hardware cost. An adapted joystick sampled by

the standard game-port served as a head movement sensor. Color monitor resolution in displaying motion was effectively enhanced 16 times by the use of anti-aliasing, enabling to display an array of 1000 random dots at real time (dual-buffer) with refresh rates of 70 Hz and above. Use of the color monitor enabled the use of the anaglyph method thus combining the manipulation of stereoscopic and monocular parallax information on a single CRT display without the loss of speed.

The potential power of our display is demonstrated here by a psychophysical measurement that was derived from the following reasoning: The subjective appearance of the depth effects derived from parallax information is very similar to the impression of depth obtained from RDS (Rogers and Graham, 1979, 1982). Moreover, there is a close similarity between the stereoscopic and monocular-parallax mechanisms, which is reflected by the nature of the tasks that are carried out by the two systems. For stereopsis, the task is to detect horizontal disparity differences between corresponding elements stimulating the two retinae simultaneously. For parallax, it is to detect differences of position or relative motion of corresponding elements stimulating the retina in two successive instances over time. So the parallax field over time for a single eye is analogous to the static disparity field for the two eyes. In viewing real 3-D objects, these two sources of information should be in concert. However, there are some essential differences between viewing real 3-D objects and a stereoscopic half image (Ritter, 1977; Lappin & Love, 1992) and egomotion contingent parallax (Ono *et al.* 1986) simulating the same object. For example, changing the viewing distance in stereoscopic display changes the perceived depth relations, while when viewing real depth, relations remain invariant in 3-D objects. The depth of the hovering segment increases with increasing viewing distance and vice versa. Another difference is apparent when you move your head parallel to a stereogram. In this case, relative shearing motion of the hovering segment, in the direction of head movements, is perceived. The percept of shearing motion (virtual parallax) is intriguing because the hovering segment is static relative to background. This study concentrated on the latter effect. First we wanted to know whether the perceived relative shift of the

central hovering segment can be nulled by introducing real but inverse motion to the hovering segment, and second, whether these compensating shears are proportional to the disparity of the hovering segment. For that study we found the Rogers and Graham technique the most suitable.

TECHNICAL REQUIREMENTS

A successful simulation of motion parallax and disparity for the study of the given task has many requirements. Motion parallax must be presented with stereoscopic cues on the same display, or rather motion parallax must be added to a random dot stereogram. For this stimulus to be presented on a raster display, the display must have a high effective resolution for displaying motion so that the apparent motion of an individual pixel jumping to one of its temporally successive neighboring pixels appears smooth and continuous. The display must be updated at a high refresh rate of at least 60 frames per second in order for the perceived motion to be smooth, and the delay of the input from the head position sensor must be sufficiently small so that no discrepancy can be detected between the head motion and the update of the stimulus.

The system that simulates the needed task and meets the criteria set above will be software based for an Intel 80386 (and beyond) microprocessor based PC compatible computer equipped with a video graphics array and color display, all standard equipment. Such systems are prevalent in research environments all over the world, and thus pose no additional cost for development: their adaptability offers short development time. Although these computers are versatile, it is difficult to meet all the requirements stated on any but the most advanced model of them. However it is possible to achieve the given goal by certain methods that in themselves successfully simulate some of the requirements.

REFRESH RATE

To achieve a high refresh rate of at least 60 frames per second on any of the systems, a low resolution video mode, 320 by 200 pixels is advisable. While the video graphics array provides

such a video mode, and it is fast compared to any of the high resolution modes, it is not fast enough when accessed directly, because on most systems video memory which is accessed through a bus interface is much slower than main memory. To overcome the problem, software-emulated double-buffering is used to create an area somewhere in the main memory that is identical to the video memory. When needed, the double-buffer in main memory is copied to video memory using processor instructions that otherwise could not be used effectively (see Appendix A). Through the use of this method there is a chance that observable video tearing can occur, as a result of an updated video memory being partially redrawn on the display due to the mismatch between updating the electron guns' position. Nonetheless, this can be corrected by synchronizing the updates of the video memory with the display's vertical retrace signal (see Appendix A), the time during which the electron guns are moving from the bottom right of the display to the top left without scanning, as this allows enough time to copy the double-buffer to video memory.

ANTI-ALIASING

Apparent continuous motion requires small changes in the position of an object. In order for this to be satisfied in the simulation a high resolution display needs to be used, but the shortage of such displays forces one to seek a software solution. The solution is best described first by a simple explanation of anti-aliasing. Anti-aliasing blends high contrast areas in an image to provide a more natural appearance and thus seemingly higher resolution than is actually available. This can be adapted to the situation faced here by assuming a higher resolution than we actually have and anti-aliasing each displayed point based on this position. Assume that each point (in the random dot stereogram) is represented on the raster display by two horizontally neighboring pixels, and we have an array of sixteen shades of a color with linear luminance that can be adapted in the display of a point. As such, a virtual horizontal resolution of 5,120 pixels can be achieved by displaying the points in the following manner. The physical horizontal position of the pixel pair will be at the virtual horizontal position modulus 16, for the 16 possible shades to use for anti-aliasing. The first

pixel in the pair will use the low 4 bits of the virtual horizontal resolution as the shade of the color, while the second pixel in the pair will be the complementary shade of the first (see Appendix B). In this manner, given a large enough viewing distance, the change of one virtual horizontal pixel is undetectable by the human eye. It should be noted that, while the horizontal resolution has been greatly enhanced, vertical resolution does not need to be stretched in this fashion, because the points only need to have horizontal translation. In practice, the left and right points of the stereogram are bitwise OR gated into the double-buffer so that duplicate pixels can be displayed correctly. For this the 256 available colors in the low resolution video mode are set so that the two shades of the colors for stereoscopic display can be represented in the upper and lower bits of the one byte that is needed to display a pixel (see Appendix C). Furthermore, to alleviate all possible delays from the display of a point, a lookup-table is created for indexing the shades of the two pixels in a point, based on their virtual horizontal position (see Appendix D).

HEAD MOTION SENSING

Head motion tracking is a key element to this simulation. The device used for head movement tracking must be able to convey position information without any observable delay. For this it was decided that the game port on the computer will be used to input the observers head position. For all practical purposes a standard analog joystick can be used where the stick can serve as a provisional chinrest. In our case, however, a 100,000 ohm 10-turn linear potentiometer was attached to the base of the head-movement sensing device. A chin rest was attached to an aluminum block with two linear bearings sliding on 9 inch parallel stainless-steel shaft providing a 20 cm track. The block was yoked with a thread to the potentiometer. The potentiometer's resistance was read by the X position pins of the game port.

The sampling rate of the game port in reading the potentiometer's resistance is dependent on the processor, its clock speed and the position of the potentiometer, because the delay is a direct result of this position. On an 486DX 50 MHz processor based system, the largest calculated delay,

allowing for cache flushes as a result of jumps, is around 80 microseconds, and approximately 600 unique positions are available on the track. Care must be taken to disable all interrupts when reading the game port, to avoid noise in the readings caused by the delay of certain interrupts firing (see Appendix E). Occasional noise that would be apparent had the interrupt flag not been cleared during a read would mostly be caused by the timer interrupt that is automatically called 18.2 times a second to update the system clock. Given that such minute procedures are not missed, the cost of production of such a device is negligible compared to 3-D head tracking systems and commercially available AD converters and is thus well suited for the simulation. The simulation as a whole is also trivial in implementation and more versatile compared to oscilloscope based systems. On the above mentioned system a maximal delay of 6 milliseconds was obtained by these software procedures (see Discussion).

SIMULATION

The presented components are combined to work in unison and display three parallel bars of random dot stereogram and thus a completed simulation. These are separated into two classes, foreground (the center stereogram) and background, and are further separated by an adjustable invisible horizontal boundary into two classes, top and bottom. Each of the classes are independent of each other and have several user adjustable settings. They can be adjusted in virtual horizontal pixel increments or physical horizontal pixel increments allowing motion from zero to plus or minus any setting, based on the observer's head movements. Similar manipulations of the disparity is possible. Colors in the disparity can be individually toggled to allow only monocular views, stereoscopic views or both. A second version of the system is capable of presenting dynamic RDS by creating several arrays of random points that are sequentially stepped through on a per frame basis. As presented here this simulation represents a Simple Integrative Method for Presenting Head Contingent Motion Parallax and Disparity Cues on Intel Processor based PCs,

developed under Turbo C/C++ and Watcom C/C++ with the use of some inline assembler language.

PRELIMINARY OBSERVATIONS

Informal observations were made with naive observers, and staff members and authors highly experienced in visual psychophysics. These observations confirmed that the system behaves satisfactorily. A random-dot pattern portraying the classical Julesz figure (a central square in front of or behind the background) presented either with parallax or disparity served as a basic stimulus. Viewing the display monocularly gave clear depth percepts that were dependent on the polarity of the differential motion. When the central square moved against the direction of the head with higher velocity it was perceived in front of the background and vice versa. The magnitude of the perceived depth was scaled by the viewing distance (Ono *et al.* 1986). Monocular depth was most vivid for small differential displacement magnitudes. The depth percept was independent of head speed and became more ambiguous at larger differential motion magnitudes (Ono *et al.* 1986). Moreover, the monocular depth effects were not as vivid as compared to the stereoscopic depth percepts as checked with the following test: The basic stimulus was split into two halves where the lower part used static disparity and the upper part used parallax. The display was observed with red-blue glasses and the lower part displayed only the red points. Observers were asked to match the binocular depth with the monocular one. Observers could not easily perform this task reliably because the monocular effect was greatly diminished under these viewing conditions. Similarly, Rogers and Collett (1989) report on 50% reduction of depth created by pure parallax (disparity was set to zero), when viewed binocularly. Finally, we checked with a few observers the virtual parallax evoked by a static RDS viewed with lateral head movements. It was reported that the magnitude of this virtual parallax increases with disparity. The latter effect was measured more systematically in the following experiment.

We note, however, that the monocular depth effects obtained with our system are not as compelling as those observed by two of us (B.J. and I.H.) in Rogers and Graham setting but were more consistent and stable then those reported by van Damme and van de Grind (1993), since our display did not show depth reversals. It seems that a more compelling monocular depth effect could be obtained by a considerable increase of the dots' density over the 1000 dots utilized (Rogers, personal communication 1996). This could be achieved in the future, without the loss of speed, by the use of a faster computer. Notwithstanding, our method shows much better dynamic performance as compared to custom low cost Head-Mounted-Tracker-Display (Forte Technologies VFX1 Headgear). The latter shows frequent noticeable discontinuities in head tracking.

EXPERIMENT

The experiment exploited the power of our method of simultaneous displaying of stereoscopic and head-movement-contingent parallax information. The task of the observers was to null virtual parallax by applying varying amounts of head-movement-contingent-parallax.

The stimulus was a 750-point random dot pattern yielding a refresh rate of 61 frames per second. The random pattern was set to portray, either by parallax or by disparity or by both, a square-wave modulation of 1.5 cycles (3 horizontal bars). The random pattern stimulus filled the whole screen (17" diagonal) subtending a visual angle of 18.9° by 15.2° at the 90 cm viewing distance. For a given measurement, the central bar was preset to have a crossed disparity from 0 to 32.2 arcmin in steps of 10.7 arcmin. The initial parallax introduced before each trial was random and ranged from ±7.2 arcmin/cm to ±14.3 arcmin/cm in 0.9 arcmin/cm increments. Random horizontal noise of ±1.8 arcmin was introduced to each point in the RDS to discourage the subjects from focusing on physical parallax between foreground and background points, aside from being given explicit instructions to observe the RDS globally. Also, a new random pattern was generated for each trial to deter the subjects from finding points that they could focus on for the parallax

nulling. Three subjects with good stereopsis, trained in psychophysics and aware of the aims of the experiment, were tested. They were given red-blue glasses to view the stereoscopic display. A method of adjustment was used throughout all the measurements. The amount of physical parallax produced to the central bar was manipulated by each individual subject, by pressing the corresponding keys on the keyboard in 0.9 arcmin/cm jumps, and could be either positive or negative, i.e. with or against the direction of the head motion, respectively. The experiment comprised of 40 random trials of the 4 disparities where no more than one of the same disparities would follow the disparity of the previous trial. All subjects reported on the perceived motion of a stationary hovering bar and had no problem in nulling this apparent parallax with the artificially introduced parallax.

RESULTS

Figure 1a, b, c depicts the results of the 3 subjects. The compensatory nulling parallax, in arcmin/cm of head movement, is plotted against the disparity of the central bar. All 3 subjects' results show a negative monotonic but decelerating decrease of the compensatory parallax. Furthermore, for all subjects and for zero disparity trials (catch trials), slight positive parallax was required. This effect is attributed to possible fatigue of disparity sensitive cells that occurred in the 30 min session (see Discussion).

DISCUSSION

A display system is presented that replicates the basic Rogers and Graham method of displaying head movement contingent parallax field added with independent binocular disparity field. The system is very adaptable and can be easily modified to include actual 3-D representation of an RDS with rotation based on the user's head movements, plotted perspectively or not. While this would require X and Y deflections of points, the anti-aliasing effect could be expanded to

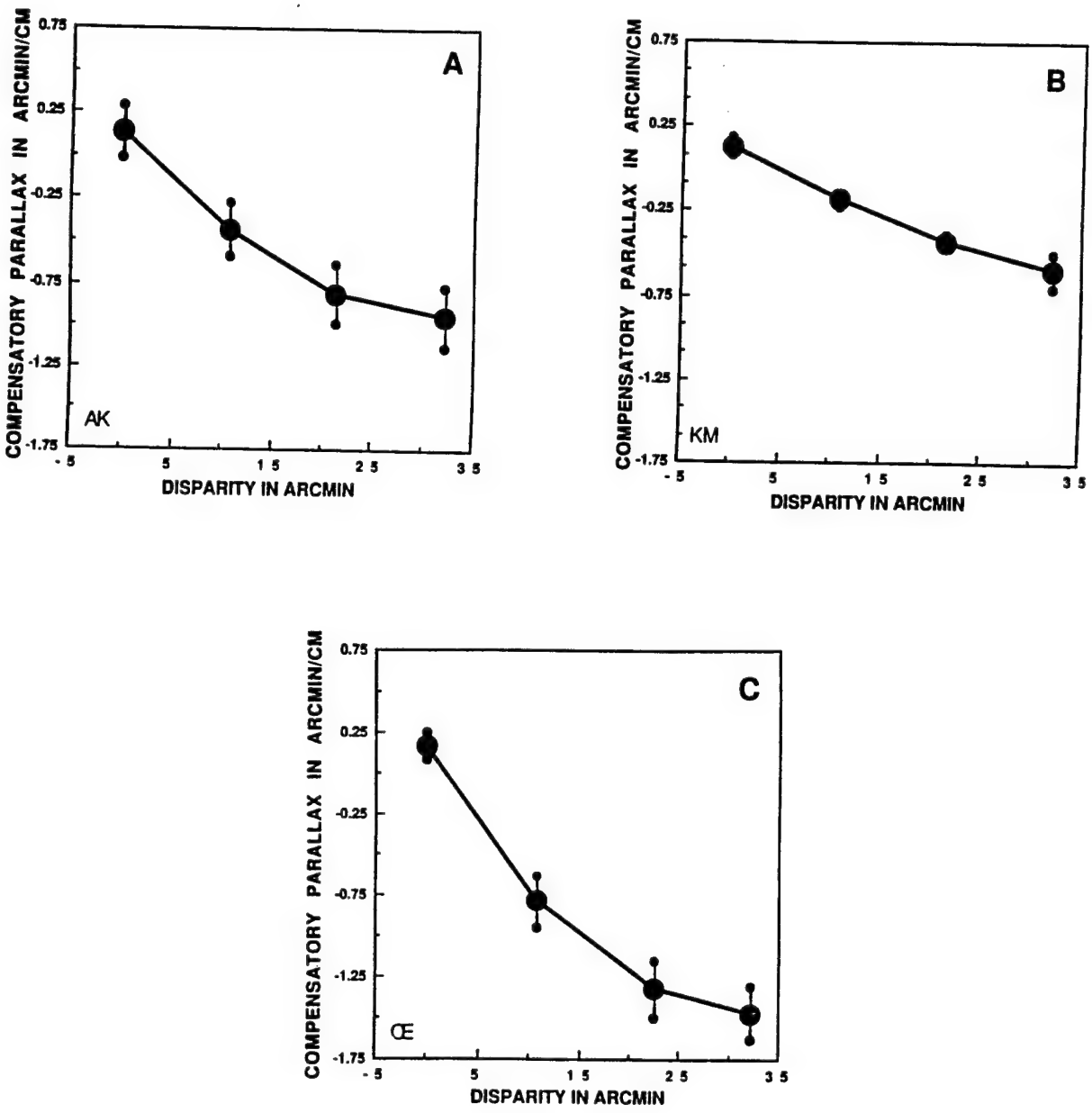


Fig. 1a, b, c

Results of 3 subjects where the compensatory annulling parallax in arcmin/cm are plotted against the stimulus disparity. Vertical bars depict one standard deviation.

allow for such motion, in monocular and stereoscopic displays. Also, with the rotation and projections of an RDS, dynamic disparity can be gained without loss in performance.

The anti-aliasing technique is required to display pixels so that subtle motion appears natural. Since a low resolution video mode is displayed, direct pixel jumps would create a jerky motion appearance at all but very fast head movements. This extra processing time needed to display an anti-aliased pixel is negligible and is very well suited for the purpose.

The current maximum delay between the start of a given sampling of the game-port and update of the screen is less than 6 milliseconds. This delay was measured by simultaneously sensing the game-port sampling pulses and the appearance of motion on the screen (sensed with a photodiode). Practically, the delay can be reduced by decreasing the number of points, increasing the speed of the computer and video card used, or both. There is a threshold after which no significant delay reduction can be gained by reduction of the number of points, and this threshold can be attributed to the time taken to swap the double-buffer. That is, when the time required to plot the points from the double-buffer is much less than the double-buffer swap time, the delay cannot be decreased significantly. This threshold for the previously mentioned processor is around 100 points and the resulting refresh rate comes to around 75 frames per second, with no vertical retrace synchronization. Furthermore, increases in processor speed and type, such as using Pentium 100 MHz processor and a faster video card, would decrease delay time considerably, to possibly around one or two milliseconds with a 1000-point RDS.

From a perceptual standpoint, the small delay inherent in our system is not significant because it is smaller than the refresh rate. Moreover, the motion of the subject's chin is linear, as dictated by the sliding chinrest. However, the head may tilt somewhat, since the forehead is not restrained by the chinrest, and may affect the motion trajectory of the subject's eyes which is the more relevant parameter. A dental bite bar method would eliminate this problem. We foresee that the future use of an advanced head-mouse prevalent in commercial virtual reality displays might

improve the temporal response of the system with the advantage of extending the head motion sensing to 3-D.

Even though the virtual parallax evoked by RDS by lateral head movements can be readily perceived, to the best of our knowledge this effect has not been investigated systematically. The system described above which combines the display of disparity as well as head movement contingent parallax enables quantitative measurement of the effect. Our results show that the virtual parallax is a monotonically increasing function of disparity, in agreement with informal observation and as confirmed by the amount of real opposite parallax that was required to null it. Interestingly, the compensatory parallax does not show a linear relationship with disparity. This is contrasted to what would be expected from pure geometrical considerations applied to real 3-D objects. In the RDS case, however, increasing the cross disparity decreased the perceived egodistance of the central bar while the background remained at the screen level. Had the stimulus been a real object, the predicted parallax would be an inverse function of egodistance, requiring increasing magnitudes of counter-parallax to annul the real parallax. Note that the compensatory parallax for all 3 subjects show small positive average value of 0.135 arcmin/cm for the zero disparity case. This is an intriguing effect since no virtual parallax should be observed in zero disparity. This effect may be attributed to the adaptation of the disparity tuned unit that were stimulated continuously during the whole experimental session. Indeed, the latter conjecture needs more empirical support.

ACKNOWLEDGEMENTS

We are indebted to Alex Kononov for his contribution in early development of the software; to Richard Payne and Patrick Grace for constructing the head movement sensor device; and to Carol A. Esso for editing and competence in word processing. This study was supported by AFOSR Grant # 93-NL-165.

REFERENCES

Ichikawa, M. and Saida, S. (1995). Interaction of binocular and motion disparity depth cues at near threshold level. Invest. Ophth. and Vis. Sci. 36, 4, 3053.

Julesz, B. (1960). Binocular depth perception of computer generated pattern. Bell System Technology Journal 39, 1126-1162.

Julesz, B. (1971). Foundations of Cyclopean Perception. University of Chicago Press, Chicago.

Lappin, J.S. and Love, S.R. (1992). Planar motion permits perception of metric structure in stereopsis. Perception and Psychophysics 51, 1, 86-102.

Ono, M.E., Rivest, J. and Ono, H. (1986). Depth perception as a function of motion parallax and absolute-distance information. Journal of Experimental Psychology 12, 3, 331-337.

Ritter, M. (1977). Effect of disparity and viewing distance on perceived depth. Perception and Psychophysics 22, 4, 400-407.

Rogers, B.J. and Collett, T.S. (1989). The appearance of surface specified by motion parallax and binocular disparity. The Quarterly Journal of Experimental Psychology 414, 4 697-717.

Rogers, B.J. and Graham, M. (1979). Motion parallax as an independent cue for depth perception. Perception 8, 125-134.

Rogers, B.J. and Graham, M. (1982). Similarities between motion parallax and stereopsis in human depth perception. Vision Research 22, 261-270.

Rogers, S.R. and Rogers, B. (1992). Visual and nonvisual information disambiguate surfaces specified by motion parallax. Perception and Psychophysics 52, 4, 446-452.

Van Damme, W.J.M. and Van De Grind, W.A. (1993). Active vision and identification of three dimensional shape. Vision Research 33, 11, 1581-1587.

APPENDIX A: VIDEO SUPPORT

In order to swap the double buffer in a timely manner, it is best if double word moves are performed so that the reading from main memory can be as effective as possible. Although the write to the video memory can only be accomplished in bytes because of hardware limitations, it is best to leave the task of breaking the double words to bytes to the processor. The following code fragments, a mix of C and assembler language, are the ones used to clear and swap the double buffer without vertical retrace synchronization.

```
int *buffer;                    declare the double buffer
int *video;                     and video buffer (which is
                                 located at A000h:0000h)
```

(clear the double buffer)

```
mov edi,[buffer]                move buffer address into 32 bit
                                 destination indexing register
mov cx,16000                    assign number of 32 bit words to
                                 counter register
xor eax,eax                     zero eax (source value)
cld                             clear the direction flag (edi is
                                 incremented)
rep stosd                       and store count words of value from
                                 eax into address from edi
```

(display the double buffer)

```
mov edi,[video]                 video into destination register
mov esi,[buffer]                buffer into source register
mov cx,16000                    count of 32 bit words
cld                             direction flag clear
rep movsd                       move
```

With retrace synchronization the double-buffer code fragment will look the same except the retrace wait code added in front, as such.

```
mov dx,0x3da                    move video port address into
                                  register dx
waitend:                        label for jump
in al,dx                        get byte from the port
test al,8                       check to see if electron guns are
                                 scanning
jnz waitend                     if not, loop until they scan
waitbegin:                      label for jump
int al,dx                       get byte from port
test al,8                       check to see if electron guns are
                                 scanning
jz waitbegin                    if yes, loop until they stop
                                 scanning and begin their retrace
mov edi,[video]                 same as before
mov esi,[buffer]
mov cx,16000
cld
rep movsd
```

APPENDIX B: ALIASED PIXEL OUTPUT

The anti-aliased point code is simple and very efficient as it uses a lookup table to find the shades for the two pixels to be drawn. There need to be two functions, one for the left and one for the right point, since the two shades for stereopsis are set so that they can be represented in either the lower or upper 4 bits of the pixel byte. In this manner the two pixels need only to be ORed into the double buffer without any preprocessing step.

```
#define ALIAS 16                define for number of positions
                                 between two pixels
int *buffer;                    double buffer declaration
```

```
int red_lookup[ALIAS];          lookup tables to return two bytes
                                 of intensities
int blue_lookup[ALIAS];         based on the intermediate
                                 (fractional) position between pixels
```

(display point with a shade of red)

```
mov eax,ebx                     ebx has the x position placed
                                 there by the compiler
and eax,15                      mask lower 4 bits to get fractional
                                 part
shl eax,2                       shift for index into 16 bit array
mov eax,[red_lookup+eax]        get 4 bytes (faster than getting 2)
shr ebx,4                       get integer part of x position
add edi,ebx                     add to destination index (which has
                                 the y position index into the
                                 buffer)
add edi,[buffer]                add the buffer address to the
                                 destination index
or [edi],ax                     write out the 2 bytes for the
                                 aliased pixels
```

(display point with a shade of blue)

```
mov eax,ebx                     same as above except using the other
                                 array
and eax,15
shl eax,2
mov eax,[blue_lookup+eax]
shr ebx,4
add edi,ebx
add edi,[buffer]
or [edi],ax
```

APPENDIX C: COLOR SETUP

The two linear luminance shades, of red and blue or red and green for stereoscopic views, can be calibrated using a luminance meter. They must be ordered so that the upper and lower bits of a display byte, and combinations of it, can contain all the possible combinations of the red and blue or red and green mixes. The following shows how the color map for the given video mode can be set to contain the correct combination of shades.

(function to set a given index in the color map to a given red, blue and green component)

```
void rgb(int c,int r,int g,int b)
{
    outp(0x3c8,c);              output color index to the index port
    outp(0x3c9,r);              and red, green & blue color values
    outp(0x3c9,g);              to the data port
    outp(0x3c9,b);              sequentially

    return;
}
```

(assuming that the array red and blue are already filled with the correctly calibrated values of the required shades, the setup of the color map is simple)

```
#define ALIAS 16

int red[ALIAS],blue[ALIAS];
int r,b;

for (b=0; b<ALIAS; b++)                         loop for blue
   for (r=0; r<ALIAS; r++)                      loop for red
      rgb(b*ALIAS+r,red[r],0,blue[b]);          blue is upper 4 bits
                                                of color index
                                                red is lower 4
```

APPENDIX D: LOOKUP TABLE INITIALIZATION

The lookup table initialization is a simple task that is performed before any points are even attempted to be displayed. The code fragment presented needs no optimization for speed, as it is only executed once every time the simulation is run. The loop below sets to two tables to their correct values

```
#define ALIAS 16

int red_lookup[ALIAS];          integer array lookup tables
int blue_lookup[ALIAS];         for pixel intensity values
int a,b,i

for (i=0; i<ALIAS; i++)
{
   a = ALIAS - i - 1;           intensity value for first pixel
   b = i;                       intensity value for second
                                 (neighboring) pixel

   red_lookup[i] = a | (b*256);          value 1 is lower 8 bits,
                                          value 2 is upper 8 bits
   blue_lookup[i] = (a | (b*256)) * ALIAS;     same as red except
                                                shifted by 4 bits
}
```

(the two multiplies by 256 are needed to left shift a given value to the following byte in a word or double word, equally the multiply by ALIAS is for the left shift of 4 bits in a byte)

APPENDIX E: INPUT HANDLING

To avoid delays, and thus noise, in reading the game port for the head position information, it is imperative that interrupts be disabled, and the loop for the read to be

accomplished as fast as possible for greater accuracy. Below is the section of code that may be used to read the game port and obtain head position information.

(the result of the read is placed in register EBX)

```
cli                         clear interrupt flag so no
                             interrupts bother us
xor ebx,ebx                 zero ebx, used for the delay count
mov dx,0x201                joystick port x deflection
out dx,al                   start the timing loop
jump:                       label
inc bx                      increment our counter
or bx,bx                    check for overflow
je done                     if overflow we abort
in al,dx                    get byte from the joystick port
and al,1                    check for cleared delay flag
jne jump                    if not clear, loop
done:                       label
sti                         allow interrupts again
```

Spin Theory & Indeterminate Scale Problem

by Alex Kononov

A thesis submitted to the

Graduate School— New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Psychology

Written under the direction of

Itzhak Hadani

and approved by

I. Hadani

Bela Julesz

Thomas V. Papathomas

New Brunswick, New Jersey

January, 1996

Spin Theory & Indeterminate Scale Problem

by Alex Kononov

A thesis submitted to the

Graduate School— New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Psychology

Written under the direction of

Itzhak Hadani

and approved by

_____

_____

_____

New Brunswick, New Jersey

January, 1996

ABSTRACT OF THE THESIS

Spin Theory & Indeterminate Scale Problem

by Alex Kononov

Thesis Director: Itzhak Hadani

A fundamental problem in the study of space perception concerns how a mobile observer acquires information about the metric structure of objects from a sequence of projections. The earlier seminal work by Hadani et al. [22] derived the optic-flow equation of a general point for an eye undergoing pure rotations. It was suggested there (without showing an explicit solution) that utilizing at least three different points and two views, the distance of object points and the motion parameters of the eye can be uniquely recovered. To substantiate these claims, current thesis shows derivation of an explicit solution for the original flow equation in the discrete case (3 points and 2 views) showing that the structure-from-motion problem can, in principle, be solved in metric terms. Additional solution is also derived in the differential approach which is applicable to a single point. Both solutions were checked with a computer simulation and have shown a remarkable degree of accuracy in a noise-free system, high sensitivity to noise of the discrete solution, and a low sensitivity to noise for the differential solution. The model was then further extended to the six degrees of freedom required to account for general motion (rotations and translations). Extended flow equations were derived which produce only partial solution. A series of additional biologically plausible constraints were, therefore, identified and used to produce analytic closed form solution.

The solutions support the Hadani et al. model which regards space perception as a navigation process where position and size constancy are preserved by reconstructing the 3-D world in terms of object's space coordinates. The significance of intrinsic metric scale in visual perception and additional biologically plausible constraints for general motion are discussed.

Acknowledgements

The author is deeply indebted to Dr. I. Hadani who made all this work possible through his unlimited patience, imagination, and most extraordinary dedication to move science beyond common sense. I am most thankful to Dr. G. Ishai and Dr. H. Frisch for providing the solid mathematical background upon which Dr. Hadani and I were able to build a mathematical structure that was previously thought to be impossible. I am most grateful to Dr. S. Klein for his comments and help in building the geometrical existence proofs for six degrees of freedom, and to Dr. B. Julesz for his encouragement. Also, this work would never be properly formated in LaTeXwithout great help from Dr. S. Marcella.

Table of Contents

Abstract ii

Acknowledgements iv

List of Tables vi

List of Figures vii

1. Introduction 1

1.1. Overview 1

1.2. Indeterminate Scale Problem 5

2. Three Degrees of Freedom: Pure Rotations 9

2.1. The Retinal Flow Equations for Pure Rotations 9

2.2. Solution of the Flow Equations in the Discrete Approach 14

2.3. Solution of the Flow Equations in the Differential Approach 15

2.4. Computer Simulations of the Two Types Solutions 16

2.5. Pure Rotations: Summary 23

3. Six Degrees of Freedom: General Motion 32

3.1. The Retinal Flow Equations for Six Degrees of Freedom 32

3.2. Solution of the Flow Equations 34

3.3. Biologically Plausible Constraints 36

3.4. Discussion 39

Appendix A. Alternative Basis 43

Appendix B. A Solution for a Single Point on a Continuous Surface . . 45

Appendix C. Flow Equations for the General Motion 47

Appendix D. General Motion, Discrete Case Along Same ϕ 49

Appendix E. Polynomial Solution for General Motion 52

Appendix F. Three Views Solution For General Motion 53

References 55

List of Tables

2.1. Errors in the predicted distance of the eye's position at different rates of measurement error, shown by the differential linear solution. 20

2.2. Errors in the predicated distance of the eye's position at different rates of measurement error, shown by the discrete nonlinear solution. . . 21

2.3. Errors in the predicated distance of the eye's position at different rates of measurement error, shown by the differential nonlinear solution. 22

List of Figures

1.1. A single spatial point projects onto the eye a Gaussian light spread function, creating a small movement field 4

2.1. Definition of systems of coordinates in the SPIN theory: X, Y, Z are fixed in space; x, y, z are attached to the eyeball. Note that the pinhole (pupil) does not coincide with the center of rotation of the eye's system. 10

2.2. Definition of retinal projection of surface points. 12

2.3. Flow chart of the simulation of the differential linear case. Unprimed notations mark true magnitudes; primed notations mark the predicted magnitudes. For details see text. 18

2.4. Stabilization of a single point by simulation of the differential linear solution for 10,000 steps. The magnified excursion of the point in the first 50 consecutive steps is shown in a circular window representing 16.8 arcmin in diameter. The polarization of the excursion steps is due to the implementation of the Listing and Donders laws in the definition of matrix M. 24

2.5. Simulation of the autokinetic movement with 0.5% colored noise penetrating from extraretinal signals. The plot depicts the excursion of a point in 100,000 steps. 25

2.6. Two possible encodings of a one-dimensional signal: a) representation by a Dirac function; b) representation by a Gaussian-like function. See text for details. 27

3.1. Definition of systems of coordinates in the SPIN theory for the general motion (rotation and translation) case. Note that the origins of X, Y, Z and x, y, z coordinate systems no longer coincide. 33

3.2. Projection of a point (P) onto a small segment of a retina: a) initial projection; b) after vertical motion and rotation (shaded); c) the same motion but distance (D) is doubled; d) after distance (D) and linear motion doubled, while rotational component is left unchanged. See text for details. 37

Chapter 1

Introduction

1.1 Overview

This thesis deals with the problem of recovering the depth (distance) of object's points and the motion parameters of the eye from the optic flow projected by a stationary environment onto a moving eye. Bruss and Horn [1] call this capacity *passive navigation* and identify several approaches used to address the issue. These can be classified into three main categories: the discrete approach (e.g., [22, 30, 29, 31, 34]), the differential approach (e.g., [10, 27]), and the least square approach (e.g., [1, 33]). Works in all three categories show that the optic flow depends upon the six motion parameters of the eye and the structure of the object. Works in the discrete and differential approach are also characterized by analyzing the minimum conditions under which an ideal observer can solve the passive navigation problem. These minimal conditions are given in terms of number of points and views. The most rigorous solution was advanced by Tsai and Huang [34]. They show that 7 points and two views are required to recover the distance and the motion parameters (up to a scalar in the translation vector). Longuet-Higgins and Prazdny [10] show that the structure of object and motion parameters can be recovered from a single point and its neighborhood (again, up to a scalar in the translation vector). The later model was criticized by Bruss and Horn [1] (also in [14]) as being incorrect except for a special case.

Excluding the Hadani et al. model, all the other models yield relative depth perception because their solution is up to a scalar. In contrast, the Hadani et al. model, now called the SPIN theory (SPIN for Space Perception In Navigation), suggests that

in addition to the eye's motion parameters, the distance of object's points from the eye can be determined uniquely from the optic flow. This is not a trivial difference because the metric (absolute) distance extraction is considered by earlier works as impossible (e.g., [36, 1, 25]). Furthermore, the SPIN theory suggests an intrinsic "hardwired" measure - the radius of the eyeball - as a metric unit for monocular depth perception. To substantiate this suggestion, this thesis presents two explicit solutions and computer simulations of the SPIN theory's optic flow equations for pure rotations [22]. The model is then extended for the general six degrees of freedom (motion consisting of rotations and translations).

To pinpoint the major difference between the earlier works and the SPIN theory approach, we concentrate on passive navigation. We raise here the question: Given that the visual system has a passive navigation capacity, what kind of perceptual experience would be predicted by the computational models when the input of the system is far below the minimal conditions required for a solution? The answer to this question is that no solution will be obtained and the system will fail to stabilize the point. From a perceptual standpoint, the appearance of the point would be jittery, reflecting the unsteadiness of the eye. It turns out that such a test is easy to administer. Take a single static point-light-source positioned 1-2m away from your eyes in a light proof room. Fixate the point when your head is fixed and turn off the lights to make the room, otherwise, dark. For most observers the point appears static for at least several seconds. Other observers may perceive it stationary for up to 30 minutes [21]. Also, you may notice that even when the point is seen starting to move, its motion shows a substantial amount of stability because the movement is smooth and slow and certainly does not reflect the unsteadiness of the eye in space which is known to exist even when we fixate. The interesting question here is why the illusory motion does not start immediately but only after a latency interval of several seconds. This question is relevant because all current models predict that the autokinetic movement (AKM) should be shown promptly at the beginning of the dark interval because a single point is below the required minimal conditions. Also, if we take as a "view" a single visual

sampling time, which is estimated to be 20-30 msec [17, 21, 2], then 10 seconds of observation, in the differential approach, may be considered as 500 successive views. Since for most observers this illusory motion is not shown for at least 10 sec (which provides enough views but not enough points) we conclude: first, that the passive navigation problem is somehow solved by the visual system even for a **single point**. Second, that all earlier models are inappropriate to account for these minimal stimulus conditions, particularly for the stability of the point in the latency interval.

To account for stability of a single point, we will show that by making a distinction between a mathematical point and a visual point, the eye's movement parameters and the distance of the point can be uniquely recovered from retinal information. All earlier works implicitly assumed that the location of a point is presented to the retina as a point intensity function (Dirac delta function) which, in principle, has no spatial extent. The reality is that the eye is a composite optical body which includes the cornea, the aqueous humor, the crystalline lens, and the vitreous humor, each of which is an imperfect optical medium. This implies that the point intensity of light is subject to refraction, diffraction, and scattering effects. Taking all these effects together, the optical impurities of the eye can be described by two-dimensional Gaussian function of light intensity, known as the physiological point spread function [35].

The point spread function used to account for the propagation of the light through the composite optical medium of the eye is assumed here to be continuous and, at least, twice differentiable. Therefore, a spatial point that may have no effective spatial extent, say a star, actually has retinal projection with spatial extent that depends on the particular cut of the point spread function with the retina. The shape of the cut (blur) depends on the light sensitivity of retinal cells and on the location of the projection on the retina. Thus, it has the shape of a circular disc near the optical axis, and of an ellipse at the periphery (Figure 1.1).

If we take the eye as a dynamic pinhole camera without a shutter and assume that the visual scene is comprised of a single luminous spatial point, then for two successive

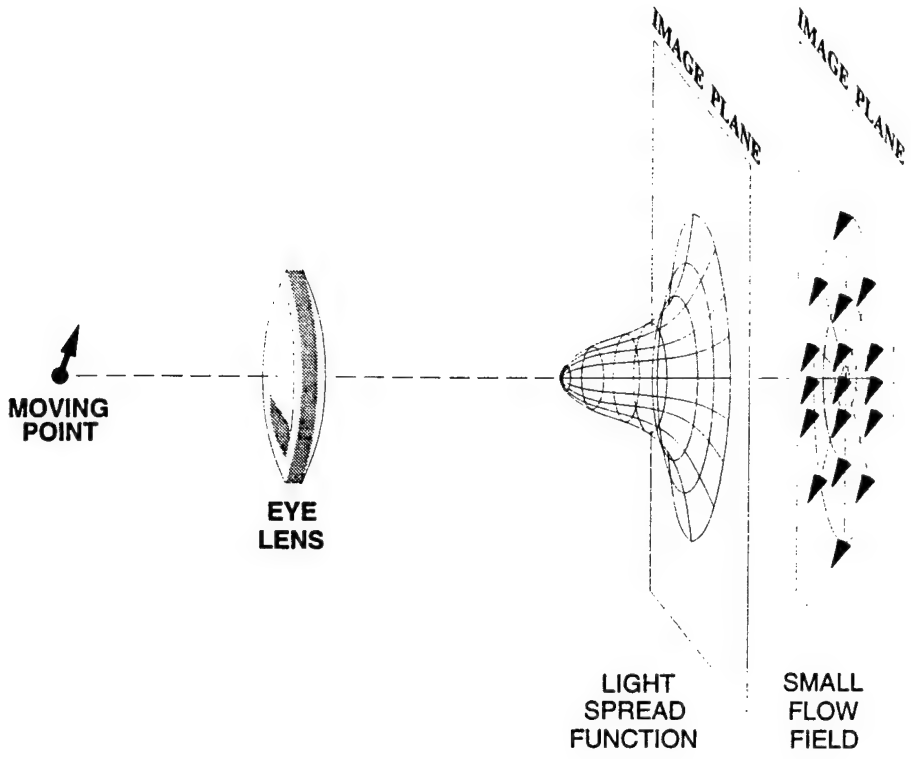


Figure 1.1: A single spatial point projects onto the eye a Gaussian light spread function, creating a small movement field

time instances t, and $t + \Delta t$, one can extract, from the positions of the two cuts of the point spread functions, a small angular velocity field at t. Based on the assumed continuity of the intensity point spread function, we further assume that the small angular velocity field is at least twice differentiable with respect to space.

Furthermore, common to most approaches mentioned above, is the use of the engineering definition for the optical flow which locates the rotation center of the camera at the pinhole [10, 1, 34]. Thus, for pure rotations, the motion field in these models is neither sensitive to the 3-D structure of object's points nor to their distance. Exceptional in this respect, is the SPIN theory approach that separates the entrance pupil (pinhole) from the center of rotation of the eye [3]. As a result, rotations of the eye (except for rotation about the optical axis, called torsion) produce for the pinhole a translation. Consequently, and in contrast to the common view, the retinal motion field is dependent on the distance and the structure of the objects even for pure eye rotations.

In this thesis we will examine the details of the indeterminate scale problem. We then describe how the problem is solved directly when motion consists of pure rotations (this material was also published in JOSA [16]). A computer simulation that was developed based on this model will then be used to predict the behavior of the system during underdetermined conditions (single point that results in AKM movement after some latency period). Next, the model will be extended to the general motion: rotations and translations. Since pure rotations are a special case of the general motion, only the differences introduced by translational components will be discussed. Finally, we will look at psychophysical implications of the SPIN model in light of passive navigation.

1.2 Indeterminate Scale Problem

Tracing the basic image formation, we observe that when a visible object is present in front of an eye it produces an image on the retina. Before looking at the details of this process and for the consistency of the description, let us adopt some formal

notation. First, let there be a three dimensional coordinate system attached to the world, such that any point in space can be described by a set of three numbers, say $[X, Y, Z]$. Next, we will need a coordinate system attached to the observer, or, more specifically, to the observer's eyeball. Any point in space, $[X, Y, Z]$, will have then a corresponding representation in observer's coordinate system, say $[X', Y', Z']$. (Unless specified otherwise, the convention adopted here will always use the *unprimed* system for the system attached to the world and the *primed* system for the system attached to the observer.) Clearly, if the exact relationship between the two coordinate systems is specified (it is known where the origin of one is located relative to the origin of the other), then a representation of a point in one coordinate system can be easily converted to the the corresponding representation in other system. Thus, for example, if we know the location of the object relative to the observer, **and** we know where observer is in space, we can compute the exact location of this object in space. Conversely, knowing observer position in space and location of some point in space, we can compute point's position relative to the observer.

The two most widely used constructs that formalize the process of image formation are *orthographic* and *perspective* projections. The simplest, the *orthographic* projection, maps a point in space to a location on the image plane (eye's retina) by projecting the point along the line parallel to the *optical axis*, or the line that is perpendicular to the retina. For the remainder of the discussion, we will adopt the convention that the observer's Y axis is the optical axis. Then, for every visible point in space, $[x', y', z']$, there is a projection on the retina, - $[x'', z'']$, where $x'' = x'$ and $z'' = z'$.

A *perspective* projection places an infinitely small pinhole along the optical axis some distance away from the retina. Then, every point in space, $[x', y', z']$, is projected onto the image plane by following a line from that point, through the pinhole, and onto an image plane, say $[x'', z'']$. It should be clear that only points along the optical axis will have $x'' = x'$ and $z'' = z'$ (same as in case of orthographic projection). Otherwise, if the pinhole is located at distance **c** from the image plane, the projection is:

$$x'' = \frac{x'c}{-y'} \tag{1.1}$$

$$z'' = \frac{z'c}{-y'} \tag{1.2}$$

A common property of either projection technique is that a three-dimensional point in space is reduced to a two-dimensional point on a plane. As a corollary, any single point on the image plane can correspond to a projection to **infinitely** many points in space. As an example, consider a set of points that have identical $[x', z']$ coordinate, but are located at a different distance from the image plane. All these points will be mapped to the exact same $[x'', z'']$ location when using orthographic projection. Similarly, by adjusting $[x', z']$ by a common factor for different distances, we get an infinite set of points with the common $[x'', z'']$ projection: $[x', y', z']$, $[2x', 2y', 2z']$, $[3x', 3y', 3z']$, etc. Therefore, it is impossible to derive a unique distance to a visible point just from the projection onto a flat plane.

A great deal of useful information can be extracted from time-varying images. At first, it might seem foolhardy to consider processing sequences of images, given the difficulty of interpreting even a single image. Curiously, though, some information is easier to obtain from a time sequence. To see this clearly, consider two images formed by a single stationary point in space from two different viewing angles. Although it is impossible to obtain the depth information from the individual projections, the two of them are enough to solve for the distance to the point if we know the exact displacement of the observer. Formally, this means that knowing the base of a triangle (displacement of the camera), and the direction angles to the point (that can be computed from the projections), we can use triangulation to estimate the distance to the point.

Two major difficulties become evident here. First, there may be no information about the observer's displacement, - a classic problem of *passive navigation* when there is no extra-retinal (active) signal present. Second, since a typical visual stimulus is not made of a one single point, but instead a huge, theoretically infinite number of visual points, identifying projection of any one individual point from view to view becomes

a non-trivial task. The latter, also known as *correspondence problem*, is addressed elsewhere [28, 4, 32, 20, 14]. Some of these references describe methods that are more applicable to finding correlation between areas (or expected futures, such as points or lines, for example) in the stereo-pairs, but replacement of spatial domain with time will make them applicable for finding correlation between views. Whichever method is used, it creates a field of brightness patters with some instantaneous velocity associated with every visible point, also called the *optical flow*. Here we will proceed under assumption that these velocities are known and will concentrate on the former issue, - the passive navigation, starting with pure rotations and then proceeding to the general case.

Chapter 2

Three Degrees of Freedom: Pure Rotations

2.1 The Retinal Flow Equations for Pure Rotations

The following sections will illustrate that, contrary to the popular view, a metric solution can in principle be obtained by a single eye when the entire motion of the observer is constrained to arbitrary rotations. The eye is modeled by a sphere, with the radius c. An infinitely small pinhole is attached to the sphere and, therefore, gives a perspective projection discussed earlier. It is important to note that, while the radius of the eyeball, c, and the optical flow (or instantaneous velocity associated with each point in the image) are known a priori, neither the distance, nor the complete characteristics of the rotational motion of the observer are given. Therefore, the two specific goals that a passive system must fulfill are: a) compute a metric distance to the visual points *and* b) compute the motion of the eye in space. The analysis of this problem starts with the analysis of the optical flow, the only information available to the passive eye about the external world.

Details of the assumption made and the logic behind the derivation of the flow equations are given elsewhere [3, 12]. Briefly, Let X, Y, Z and $x(t)$, $y(t)$, $z(t)$ be two Cartesian coordinate systems with a common origin coinciding with the center of rotation of the eyeball, as shown in Figure 2.1. X, Y, Z is fixed to the head (and since, by assumption, the head is fixed in space, this system is also fixed in space). Let $x(t)$, $y(t)$, $z(t)$ be a moving system attached to the eyeball where the y axis coincides with the optical axis. Each static object point $R = [X, Y, Z]$ can be expressed in terms of

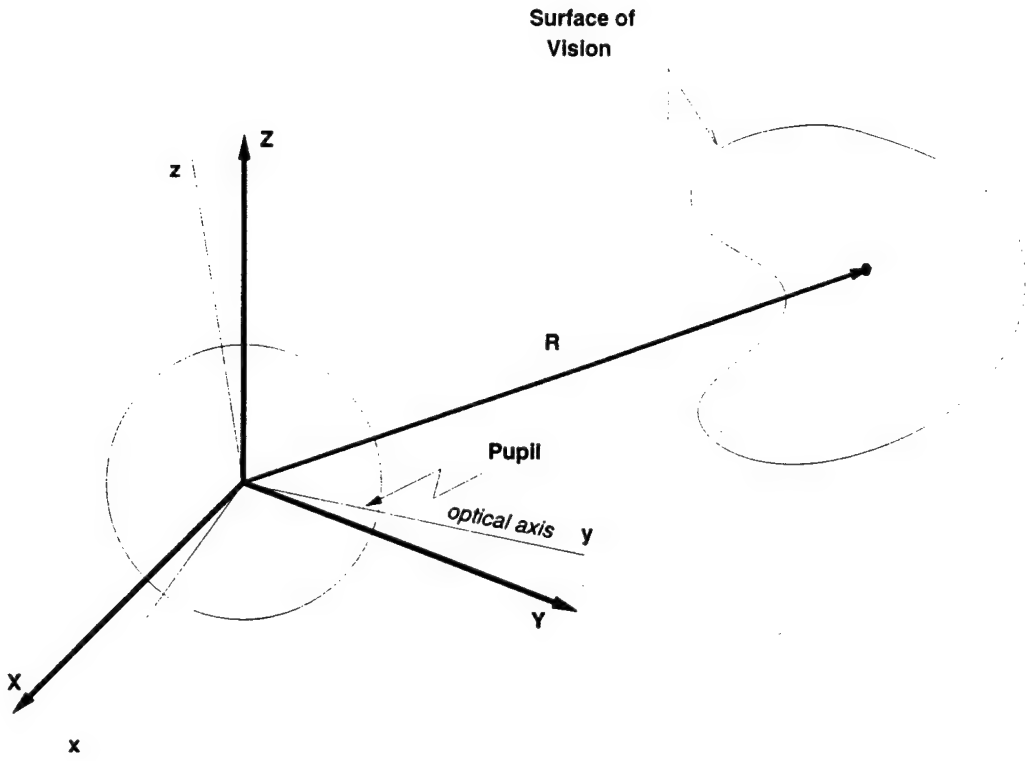


Figure 2.1: Definition of systems of coordinates in the SPIN theory: X, Y, Z are fixed in space; x, y, z are attached to the eyeball. Note that the pinhole (pupil) does not coincide with the center of rotation of the eye's system.

the eye system coordinates $\mathbf{r}(t) = [x(t), y(t), z(t)]$ by the linear transformation

$$\mathbf{r}(t) = \mathbf{M}(t)\mathbf{R} \tag{2.1}$$

where $\mathbf{M}(t)$ is 3×3 orthogonal transformation matrix ($\mathbf{M}^T * \mathbf{M} = \mathbf{I}$). Having this mapping in mind x(t), y(t), z(t) can additionally be mapped onto retinal polar coordinates (Figure 2.2) which are also given in terms of the eye system by

$$\theta = \frac{2cos^{-1}y - c}{\sqrt{x^2 + (y-c)^2 + z^2}} \tag{2.2}$$

$$\phi = tan^{-1}(\frac{x}{z}) \tag{2.3}$$

where c is the radius of the eyeball. A third coordinate ρ is defined as the distance between the pinhole and the location of the object, where:

$$\rho = \sqrt{x^2 + (y-c)^2 + z^2}. \tag{2.4}$$

Using equations (2.2)-(2.4), the inverse, mapping of the retinal projection points onto surface points is given by

$$x = \rho sin\frac{\theta}{2} sin\phi \tag{2.5}$$

$$y = c + \rho cos\frac{\theta}{2} \tag{2.6}$$

$$z = \rho sin\frac{\theta}{2} cos\phi \tag{2.7}$$

Note that the transformations (2.2)-(2.4) and (2.5)-(2.7) are not time dependent because they are carried out within the eye's system.

The elements of the matrices $\mathbf{M}(t)$ are functions of the directional angles of the eye in the socket. In principle, three independent parameters $\lambda(t)$, $\mu(t)$, $\nu(t)$ are required to define the relative position of eye and head system. However, Listing and Donders' laws of eye movements mean that actually two parameters $\lambda(t)$ and $\mu(t)$ are utilized by the visual system [23]. For the present solution we adopted this interpretation of the

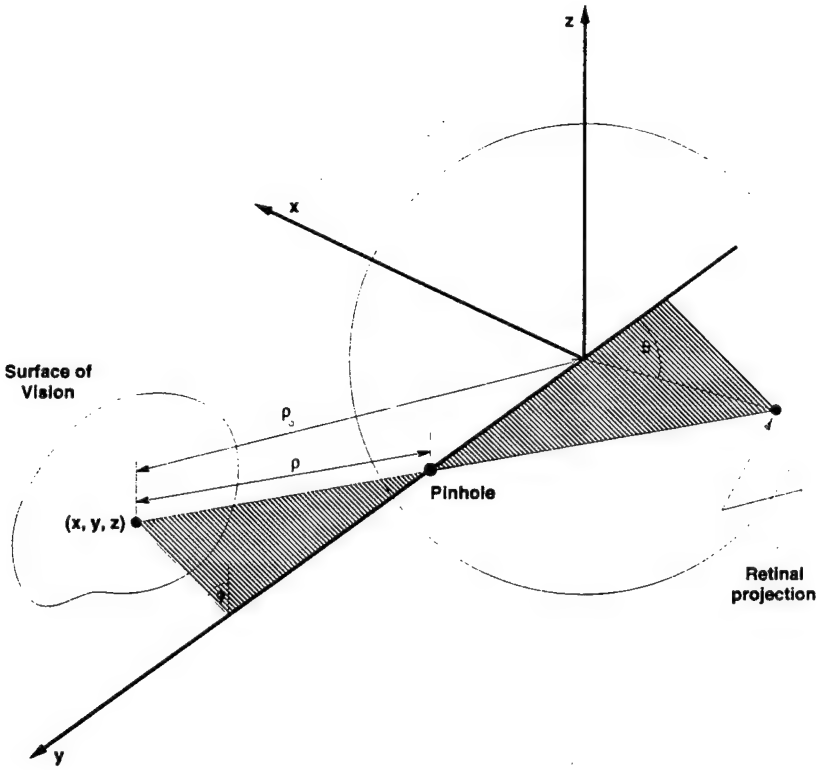


Figure 2.2: Definition of retinal projection of surface points.

x,y,z	are the coordinates of a spatial point transformed into the eye's notation;
θ,ϕ	are eccentricity and meridian angles, respectively;
ρ,ρ_0	are distances of a surface point from the pinhole and from the origin, respectively;
c	is the radius of the eyeball.

two laws and used the following matrix M:

$$\mathbf{M} = \begin{bmatrix} cos^2\mu cos\lambda + sin^2\mu & -cos\mu sin\lambda & cos\mu sin\mu(cos\lambda - 1) \\ sin\lambda cos\mu & cos\lambda & sin\mu sin\lambda \\ sin\mu cos\mu(cos\lambda - 1) & -sin\mu sin\lambda & sin^2\mu cos\lambda + cos^2\mu \end{bmatrix} \tag{2.8}$$

where λ and μ are the two directional angles. Since the object's points are stationary, the absolute derivative of $\mathbf{r}$ with respect to time is:

$$\frac{d\mathbf{r}}{dt} = \dot{\mathbf{r}} + \omega \times \mathbf{r} = 0 \tag{2.9}$$

where $\dot{\mathbf{r}}$ is the time derivative of $\mathbf{r}$ in terms of the eye system, and ω is the angular velocity vector of the eye given in terms of the eye-system notations. From equation (2.9) we obtain:

$$\dot{x} + \omega_y z - \omega_z y = 0 \tag{2.10}$$

$$\dot{y} - \omega_x z + \omega_z x = 0 \tag{2.11}$$

$$\dot{z} + \omega_x y - \omega_y x = 0 \tag{2.12}$$

Differentiating (2.5)-(2.7) with respect to time, substituting for $\dot{x}$, $\dot{y}$, $\dot{z}$ in (2.10)-(2.12) and eliminating $\dot{\rho}$ [3] we get

$$\dot{\theta} = \left(2 + \frac{2ccos\frac{\theta}{2}}{\rho}\right)(\omega_z sin\phi - \omega_x cos\phi), \tag{2.13}$$

$$\dot{\phi} = \left(\frac{\rho cos\frac{\theta}{2} + c}{\rho sin\theta_2}\right)(\omega_x sin\phi + \omega_z cos\phi) - \omega_y. \tag{2.14}$$

These flow equations establish the relationships between the components of the retinal movement vector $\dot{\phi}$, $\dot{\theta}$ and: a) the three components of the angular velocity vector ω, and b) the distance ρ of the object point. The generality of the equations stems from the fact that these analytic relationships hold at any retinal location. The magnitude c which represents the radius of the eyeball is taken as the metric unit of the solution.

2.2 Solution of the Flow Equations in the Discrete Approach

Equations (2.13) and (2.14) are now solved for the discrete case. We first define two new parameters A and B as:

$$A = \omega_x sin\phi + \omega_z cos\phi, \tag{2.15}$$

$$B = \omega_z sin\phi - \omega_x cos\phi. \tag{2.16}$$

We substitute A and B in (2.13) and (2.14). Then, we isolate c/ρ in each of these equations and equate both expressions for c/ρ. These manipulations yield the following single equation in A, B and ω_y:

$$\omega_y = \dot{\theta}\frac{A}{B}\frac{1}{sin\theta} - Atan\frac{\theta}{2} - \dot{\phi} \tag{2.17}$$

The ratio $\frac{A}{B}$ is:

$$\frac{A}{B} = \frac{\omega_x sin\phi + \omega_z cos\phi}{\omega_z sin\phi - \omega_x cos\phi} = \frac{tan\phi + \frac{\omega_x}{\omega_z}}{\frac{\omega_x}{\omega_z}tan\phi - 1}. \tag{2.18}$$

Substituting A and $\frac{A}{B}$ in (2.17) yield for any given point i:

$$\omega_y = \frac{(tan\phi_i + \frac{\omega_x}{\omega_z})\dot{\theta}_i}{(\frac{\omega_x}{\omega_z}tan\phi_i - 1)sin\theta_i} - \omega_x tan\frac{\theta_i}{2}(sin\phi_i + cos\phi_i\frac{\omega_z}{\omega_x}) - \dot{\phi}_i, \tag{2.19}$$

$$i = 1, 2, \cdots, N.$$

Equation (2.19) comprises a set of N equations in 3 unknowns, ω_x, ω_y, ω_z, but is given in terms of the two parameters $\frac{\omega_x}{\omega_z}$ and ω_x. Since ω_y is a common unknown, the right side of equation (2.19) holds for any image point. If we consider two such points we can write an implicit equation in $\frac{\omega_x}{\omega_z}$ for any two different points i, j:

$$\omega_x = \frac{\dot{\phi}_j - \dot{\phi}_i + \frac{\dot{\theta}_i(tan\phi_i + \frac{\omega_x}{\omega_z})}{sin\theta_i(\frac{\omega_x}{\omega_z}tan\phi_i - 1)} - \frac{\dot{\theta}_j(tan\phi_j + \frac{\omega_x}{\omega_z})}{sin\theta_j(\frac{\omega_x}{\omega_z}tan\phi_j - 1)}}{(sin\phi_i + \frac{\omega_x}{\omega_z}cos\phi_i)tan\frac{\theta_i}{2} - (sin\phi_j + \frac{\omega_x}{\omega_z}cos\phi_j)tan\frac{\theta_j}{2}} \tag{2.20}$$

$$i = 1, 2, \cdots, N; j = 1, 2, \cdots, M; N \neq M.$$

Since ω_x is also a common unknown, equation (2.20) can be further written as a single equation for three different points i, j, k with the ratio $\frac{\omega_x}{\omega_z}$ as unknown. This yields a fourth order polynomial in $\frac{\omega_x}{\omega_z}$. Solving the polynomial yields 4 roots. The root that,

when used to calculate ρ, gives a maximum positive value is the correct solution. The computer simulation described below has shown that this solution is very sensitive to noise measurement. To overcome this difficulty we now turn to a linear solution of the flow equations for a single visual point.

2.3 Solution of the Flow Equations in the Differential Approach

To obtain a solution for a single point we assume that the surface of objects of vision is composed of densely packed discrete points which have no extension. Since each individual point, by definition, has no structure, the light emanating from this point, projects on the retina due to the optical impurities of the eye, a Gaussian intensity function that has a spatial extension. Thus, when the point starts to move across the retina due to eye movements, a small movement field is created which eventually can be extracted by retinal and brain mechanisms [3, 20]. To derive an analytic solution for that small movement field, we first substitute A and B into equations (2.13) and (2.14) and differentiate the result with respect to θ and ϕ under the assumption $\rho_\theta = \rho_\phi = 0$. This produces:

$$\dot{\theta}_\theta = -\frac{c}{\rho} B sin\frac{\theta}{2}, \tag{2.21}$$

$$\dot{\theta}_\phi = 2A\left(1 + \frac{c}{\rho} cos\frac{\theta}{2}\right), \tag{2.22}$$

$$\dot{\phi}_\theta = -\frac{A}{2sin^2\frac{\theta}{2}}\left(1 + \frac{c}{\rho} cos\frac{\theta}{2}\right), \tag{2.23}$$

$$\dot{\phi}_\phi = -\frac{B}{sin\frac{\theta}{2}}\left(cos\frac{\theta}{2} + \frac{c}{\rho}\right). \tag{2.24}$$

Equations (2.13)-(2.14) and (2.21)-(2.24) comprise a set of 6 independent equations that provide more than is required to solve for the three unknowns ρ, A, and B. Furthermore, by taking the second spatial derivative of $\dot{\phi}$ with respect to ϕ, we arrive at the following equation for the fourth unknown:

$$-\omega_y = \dot{\phi} + \dot{\phi}_{\phi\phi}. \tag{2.25}$$

Explicit expressions for the other three unknowns, given in terms of the observable quantities, are:

$$\rho = -\frac{c\dot{\theta}}{2\dot{\theta}_\theta} sin\frac{\theta}{2} - ccos\frac{\theta}{2}, \tag{2.26}$$

$$A = \frac{\dot{\theta}_\phi}{2} + \frac{\dot{\theta}_\phi \dot{\theta}_\theta}{\dot{\theta}} cot\frac{\theta}{2}, \tag{2.27}$$

$$B = \frac{\dot{\theta}}{2} + \dot{\theta}_\theta cot\frac{\theta}{2}. \tag{2.28}$$

Solving for A and B, ω_x and ω_z can be easily recovered from (2.15) and (2.16).

Equations (2.25)-(2.28) establish a set of 4 linear equations which have a unique solution. Indeed, these equations were derived under the assumption that the individual object's point has no structure. Thus, we could set to zero the two unknown quantities ρ_θ, ρ_ϕ that otherwise would appear in (2.21)-(2.24). This assumption is entirely plausible when we consider points that are relatively far from the pinhole (stars) but yields an approximate solution when the points are at a close distance from the pinhole. However, in *Appendix B* we show a more general solution that can solve the problem for close points which have a structure, or alternatively, when the object of vision is a continuous surface.

2.4 Computer Simulations of the Two Types Solutions

The computer simulations focused on the kinematical aspects of the passive navigation process. For each small rotation of the eyeball 4 unknown magnitudes were calculated: The 3 components of the eye's angular velocity vector ω, and the distance(s) of object points. These unknowns were derived on the basis of the observable quantities $\dot{\theta}$, $\dot{\phi}$ and their spatial derivatives.

In order to carry out the simulations for prolonged durations which are composed of many consecutive small random steps, we derived the relations between the components of ω that are calculated from the optic flow and the momentary orientation of the eye relative to the head. The latter is given by the two position angles λ, μ. This is done

by solving the differential equation:

$$\dot{\mathbf{B}} = \Omega\mathbf{B}, \tag{2.29}$$

where

$$\Omega = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \tag{2.30}$$

Utilizing a transformation matrix $\mathbf{M}(t)$ that is given in terms of two position angles means that this aspect of the simulation emulated the Listing and Donders' laws of eye movements. One meaning of these laws is that ω_y is a function of ω_x, ω_z, λ and μ. We note, however, that to simulate a pure passive navigation, ω_y can be computed directly from equations (2.19) or (2.25).

Simulations of the discrete and the differential cases were written in C language and run on a SUN Sparc Station 2. A flow chart of the simulation of the differential linear case is shown in Figure 2.3. After the input stage $i = 0$, every step has several stages as described below:

Input. Feeding the computer with the spatial coordinate of the point(s) and the components of the eye's angular velocity vector (first step only).

Projection onto the retina of spatial point(s). The spatial coordinates of point(s) are transformed by the (true) matrix $\mathbf{M}$ to the momentary eye's system Cartesian components. The latter are then transformed to their retinal polar equivalent notation. The components of the true rotation vector ω embedded in A and B, are used to update the elements of the transformation matrix $\mathbf{M}$ to be used in the subsequent step.

Processing of the retinal information. The optical flow equations are solved to obtain the predicted ρ' and ω'. In turn, these are used to update the elements of the (predicted) matrix $\mathbf{M}'$.

Projection out in space. The predicted distance ρ' and the retinal polar coordinates θ and ϕ are transformed to the eye's Cartesian coordinates. The later are "projected-out" to their predicted spatial coordinates by applying the inverse of the matrix $\mathbf{M}'$.

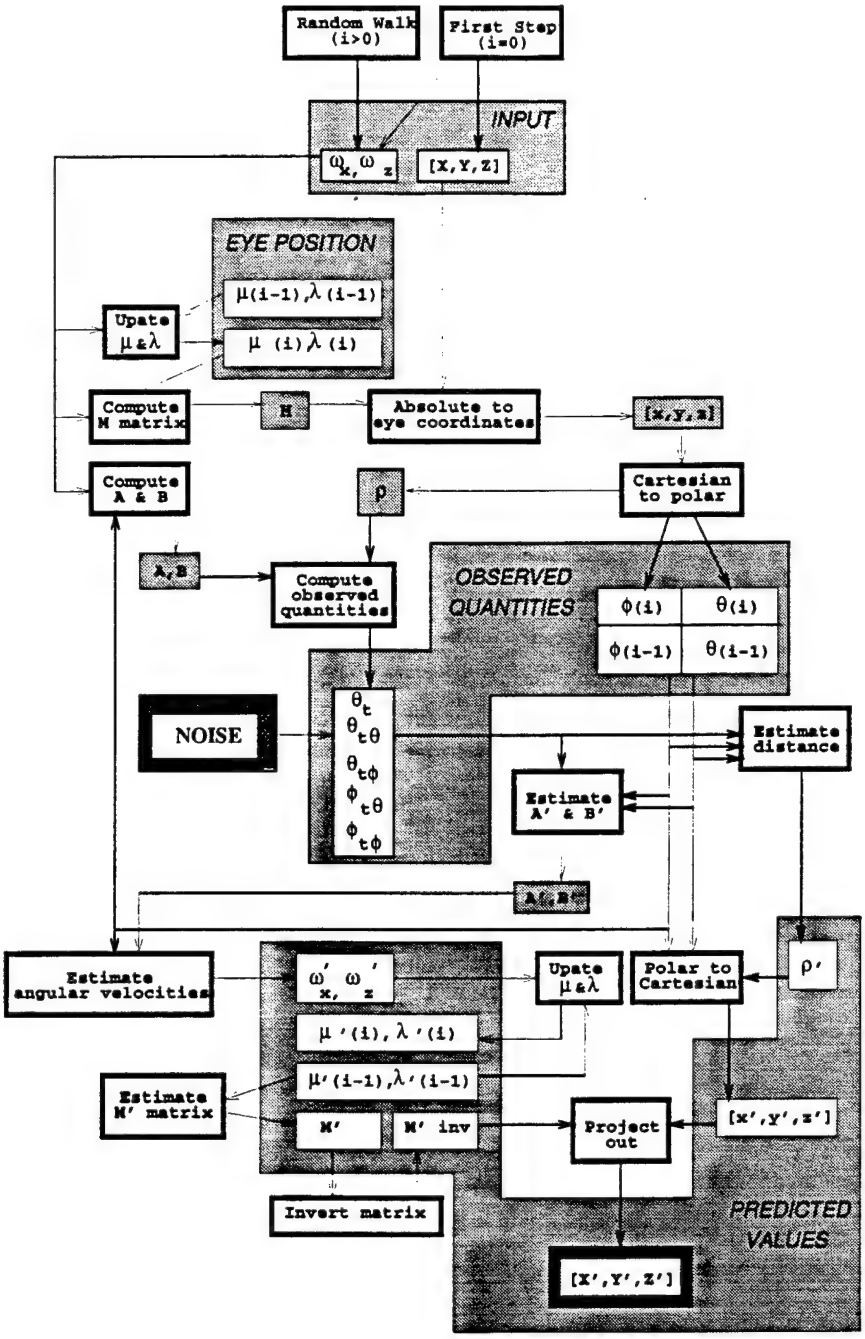


Figure 2.3: Flow chart of the simulation of the differential linear case. Unprimed notations mark true magnitudes; primed notations mark the predicted magnitudes. For details see text.

Thereafter, the computer generates a new random rotation to the eyeball.

Note that in the process of 3-D reconstruction, the algorithm does not end up only with two consecutive frames, but involves many small random steps. These random steps are used to calculate the new position of the eyeball by updating the elements of $\mathbf{M}$ and $\mathbf{M}'$. Note that $\mathbf{M}$ represents the actual position of eye in the socket and is used to project the point(s) onto the retina. The inverse of $\mathbf{M}'$ is used to project the point out in space. Thus, small discrepancies between the true and the predicted values, in any given step, are integrated across many consecutive steps to create a large orientation error in $\mathbf{M}'$. This may lead to the accumulation of orientation error in the projection-out phase that results in apparent movement of the reconstructed static point. In this respect, the simulation emulates the SPIN theory's conception about the nature of the autokinetic effect (see below).

All types of the above simulations were run for up to 300,000 consecutive steps to check for: a) the correctness of the mathematical derivations, b) to determine the baseline of the computational error in a noise-free system, and c) to evaluate and compare the noise immunity of the different algorithms. Note that the use of small rotations emulates the small excursions of the eyeball that occur during fixation. These include slow drift, tremor and microsaccade [1]. They certainly do not simulate large saccadic eye movements (in this case the elements of $\mathbf{M}'$ are assumed to be updated by extraretinal signals). Note also that each step is roughly equivalent to 20 msec of fixation time, and 3600 steps are equivalent to one minute of fixation.

The accuracy of all simulations in a noise-free system was remarkably high. Even after 300,000 steps the computational error of all the predicted parameters did not exceed 0.001error has shown that the most robust solution is the differential linear (Table 2.1). Maximum errors in the calculated distance were proportional to the measurement error at all distances (up to 1 Km). This high immunity to noise is expected because the set of equations is linear. However, the simulations of the nonlinear discrete solution

Table 2.1: Errors in the predicted distance of the eye's position at different rates of measurement error, shown by the differential linear solution.

Level Noise in $\dot{\theta}$ (%)	ρ (cm)	Average Error ρ' (%)	Maximum Error μ' (deg)	Maximum Error λ' (deg)
1	100,000	1.00	0.08	0.01
	10,000	1.00	0.05	0.03
	1,000	1.00	0.06	0.14
	150	1.01	0.15	0.04
	75	1.01	0.11	0.05
	35	1.02	0.19	0.04
	18	1.03	0.06	0.06
	9	1.02	0.12	0.03
5	100,000	4.99	0.50	0.10
	10,000	5.00	0.24	0.17
	1,000	5.00	0.28	0.14
	150	5.03	0.70	0.25
	75	5.06	0.26	0.15
	35	5.11	0.52	0.28
	18	5.13	0.79	0.21
	9	5.08	0.53	0.24
10	100,000	10.00	0.40	0.19
	10,000	9.99	0.52	0.44
	1,000	10.01	0.43	0.84
	150	10.06	0.67	0.61
	75	10.12	1.14	0.46
	35	10.21	1.81	0.53
	18	10.26	2.10	0.29
	9	10.18	1.97	0.52
20	100,000	20.00	1.30	0.34
	10,000	19.99	0.79	0.24
	1,000	20.02	1.28	1.64
	150	20.12	3.04	0.68
	75	20.24	1.51	1.27
	35	20.44	3.12	1.39
	18	20.51	1.68	0.80
	9	20.42	1.44	0.89

Note: Data in each row are based on 3,000 steps per run.

Table 2.2: Errors in the predicated distance of the eye's position at different rates of measurement error, shown by the discrete nonlinear solution.

Noise Level in $\dot{\theta}$ (%)	ρ (cm)	Maximum Error ρ (%)	Average Error ρ' (%)	Maximum Error μ' (deg)	Maximum Error λ' (deg)
10^{-4}	150	24.04	3.06	0.01	0.00
	75	5.32	0.83	0.01	0.00
	35	2.79	0.17	0.00	0.01
	18	1.20	0.06	0.00	0.00
	9	0.29	0.03	0.00	0.00
10^{-3}	150	-	-	-	-
	75	-	-	-	-
	35	79.53	2.22	0.03	0.00
	18	7.15	0.56	0.04	0.00
	9	14.44	0.42	0.00	0.00
10^{-2}	150	-	-	-	-
	75	-	-	-	-
	35	-	-	-	-
	18	-	-	-	-
	9	-	-	-	-
10^{-2}	35	-	-	-	-
	18	-	-	-	-
	9	-	-	-	-
10^{-2}	18	-	-	-	-
	9	-	-	-	-

Note: Data in each row are based on 3,000 steps per run. Dashes represent unacceptable results.

Table 2.3: Errors in the predicated distance of the eye's position at different rates of measurement error, shown by the differential nonlinear solution.

Noise Level in $\dot{\theta}$ (%)	ρ (cm)	Maximum Error ρ (%)	Average Error ρ' (%)	Maximum Error μ' (deg)	Maximum Error λ' (deg)
10^{-4}	150	0.79	0.11	0.00	0.00
	75	0.16	0.03	0.00	0.00
	35	0.02	0.01	0.02	0.00
	18	0.00	0.00	0.10	0.01
	9	0.00	0.00	0.23	0.01
10^{-3}	150	8.75	1.18	0.01	0.00
	75	1.26	0.29	0.00	0.00
	35	0.17	0.05	0.03	0.00
	18	0.04	0.01	0.10	0.01
	9	0.01	0.01	0.24	0.01
10^{-2}	150	266.37	13.78	0.10	0.01
	75	16.72	2.95	0.04	0.00
	35	1.77	0.56	0.03	0.00
	18	0.35	0.13	0.10	0.01
	9	0.10	0.04	0.25	0.01
10^{-2}	35	21.23	5.64	0.11	0.02
	18	3.90	1.31	0.11	0.01
	9	1.00	0.36	0.23	0.01
10^{-2}	18	38.28	13.71	0.41	0.06
	9	11.30	3.56	0.24	0.10

Note: Data in each row are based on 3,000 steps per run.

and nonlinear differential (given in the *Appendix B*), show a high sensitivity to measurement error (Tables 2.2 and 2.3). The sensitivity to noise increased with increasing the distance of the points from the pinhole. Reasonable results with the discrete nonlinear case could be obtained up to distance of 150cm and with a measurement error of 10-4discrete case and gave reasonable results at 150cm distance with noise rates below 10-2is considered, distance estimations calculated by the two nonlinear solutions will be considerably improved because the optic flow is more sensitive to parallax for eye's translations than for eye's rotations.

The high degree of accuracy obtained in the linear solution can account for the latency interval of the autokinetic movement, e.g. for the time interval where a point appears static. Introducing measurement error resulted in polarized jittery fluctuations of the reconstructed coordinates of the point (Figure 2.4). This effect is shown because in the definition of the transformation matrix **M**, we implemented the Listing and Donders' laws.

As shown in Tables 2.1 and 2.2, the magnitude that is most sensitive to measurement error in all types of solutions is ρ', while errors in μ' and λ' remain relatively small. This is because of the uniform distribution noise utilized. This type of noise averages to zero and therefore cannot produce large autokinetic movements. On the basis of these results we concluded: First, that a non-uniform noise function may be required. Second, the noise should be introduced directly to the components of ω emulating noise that penetrates from oculomotor signals. An example of a simulation with "extraretinal noise" is shown in Figure 2.5. It was obtained by introducing colored noise (a white noise that is integrated by first order linear system) directly to the components of ω'.

2.5 Pure Rotations: Summary

In this work explicit solutions for the passive navigation problem in pure rotations were derived and tested. Two types of solutions were presented for the optic flow equations suggested by the SPIN theory. The first is given in the discrete approach

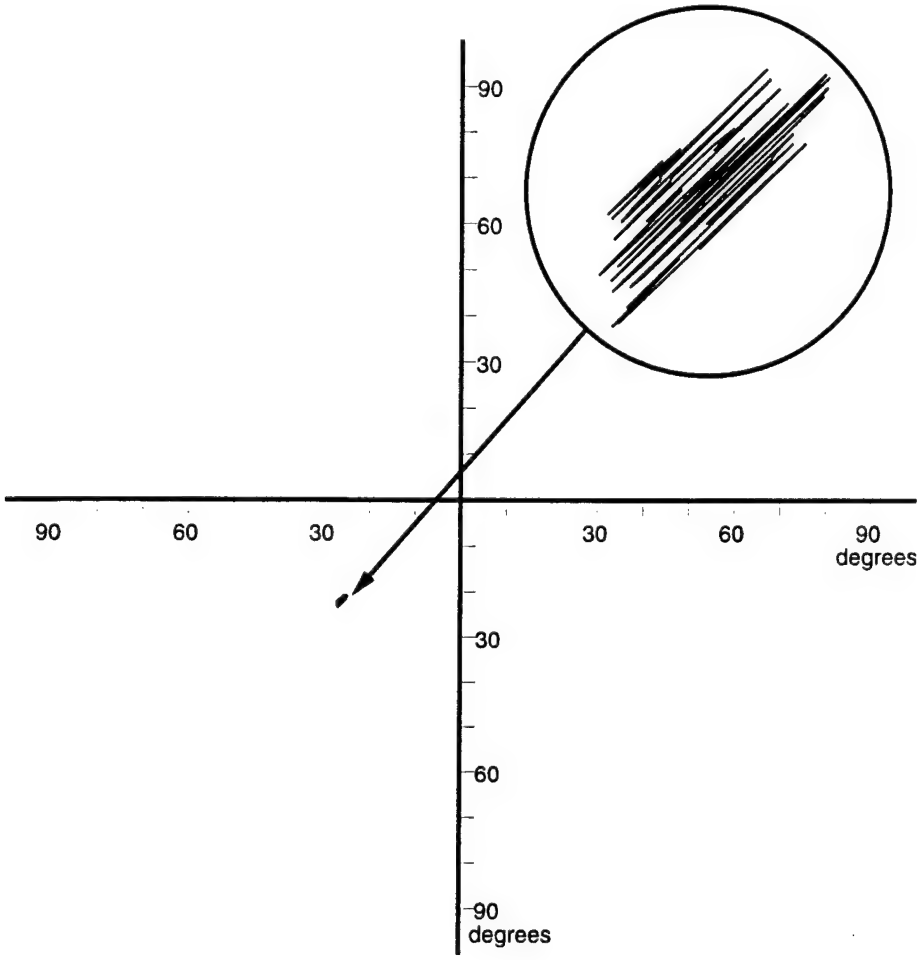


Figure 2.4: Stabilization of a single point by simulation of the differential linear solution for 10,000 steps. The magnified excursion of the point in the first 50 consecutive steps is shown in a circular window representing 16.8 arcmin in diameter. The polarization of the excursion steps is due to the implementation of the Listing and Donders laws in the definition of matrix M.

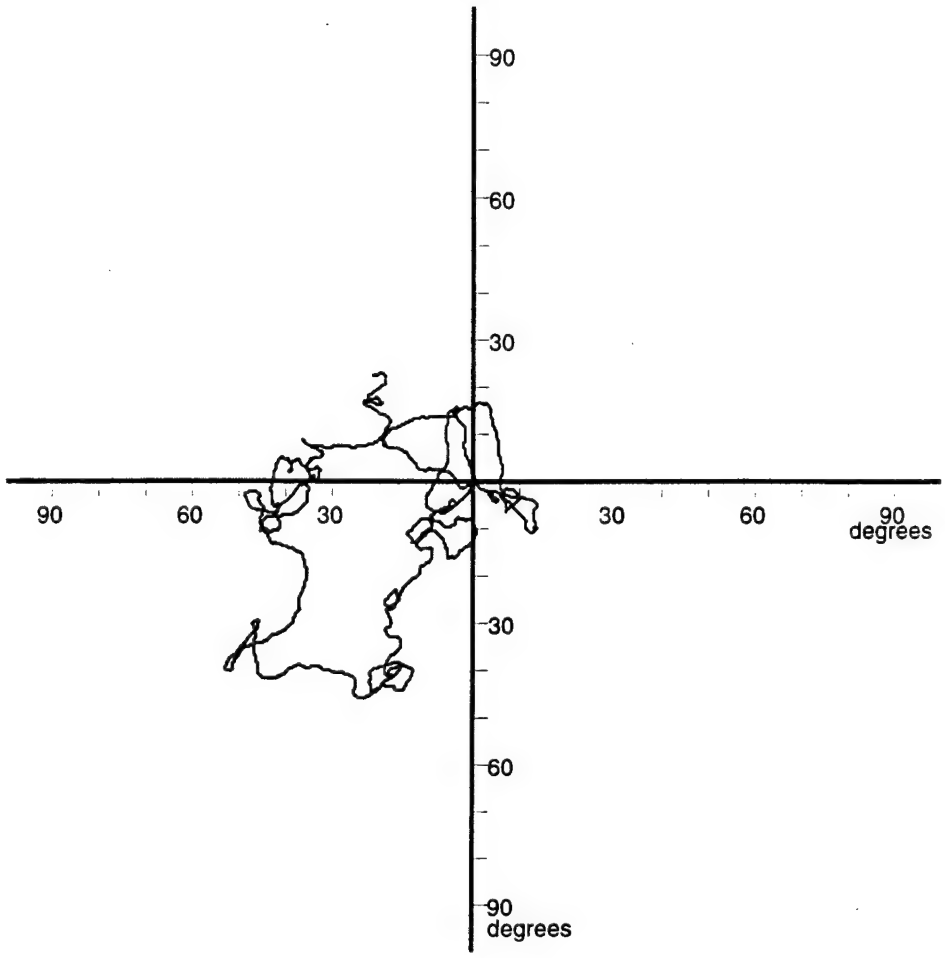


Figure 2.5: Simulation of the autokinetic movement with 0.5% colored noise penetrating from extraretinal signals. The plot depicts the excursion of a point in 100,000 steps.

and requires three points and two views (or time derivative). The second is given in the differential approach and is applicable to a single point. The first solution does not involve any assumptions other than those made in Hadani et al. paper [3], e.g. that the visual scene is comprised of static objects, and the pinhole is separated from the eye's center of rotation. Thus, this solution just substantiates the statement made in that paper saying that information provided by additional points is not redundant. However, this solution requires a system that can solve a set of non-linear equations. This, in turn, requires solving a fourth order polynomial ending up with 4 roots or 4 alternative solutions. Even though the criterion to select the correct root is well defined, the solution is not unique. Moreover, the simulations found this solution to be the most sensitive to measurement error. These drawbacks were removed in the second type of solution which is given in the differential approach. This approach involves the solution of only 4 linear equations and therefore the solution is unique. Apart from being simpler and unique, this solution is applicable for minimal visual conditions comprised of a single point. Both solutions show that the Cartesian space coordinates of object's points and the movement parameters of the eye can, in principle, be recovered from the time derivative of the optic flow. Furthermore, both are metric solutions in which the metric unit is the radius of the eyeball. While the implications of the present results to machine vision cannot be overlooked, we restrict the discussion to the relevance of the procedures to human vision because the motivation was mainly inspired by the features of perceptual systems rather than by engineering systems.

The solutions in the differential approach are novel and may be considered as the main contribution of the present work. They are novel because they are based on a new assumption, e.g that the retinal projection of any visible spatial point has an extension. While the plausibility of this assumption can hardly be refuted, it is interesting to note that a system can utilize what is normally considered as noise, or blur, to solve an indeterminate problem for a single point.

Actually, there is a more fruitful way to look at this phenomena. Since the retina is composed of individual sensory elements (cones and rods), it can be functionally

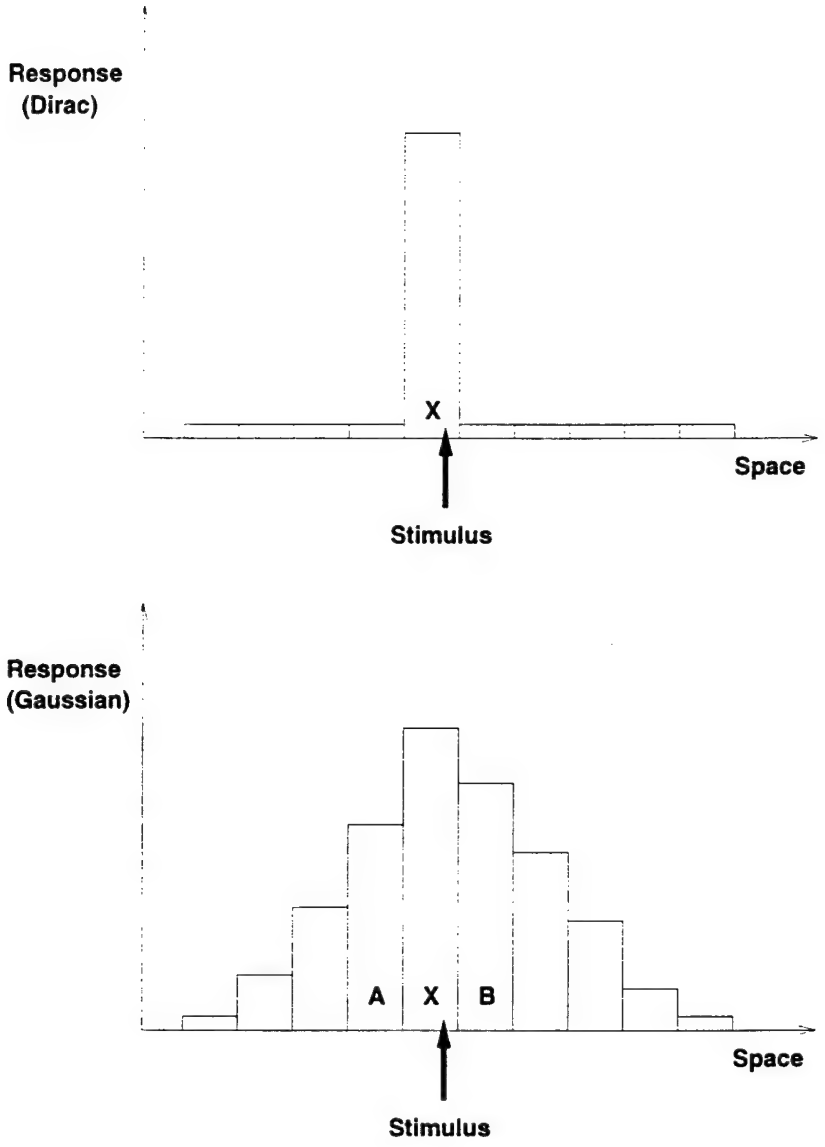


Figure 2.6: Two possible encodings of a one-dimensional signal: a) representation by a Dirac function; b) representation by a Gaussian-like function. See text for details.

considered a *digital* system. Thus, for example, while another biological digital system, cochlea, was proposed by Helmholtz [38] to encode sound by producing a narrow response for each specific frequency (resonance theory), it was later refuted by the experiments conducted by Georg von Bekesy in the 1920s and 1930s. Von Bekesy found [37], in contradiction to the resonance hypothesis, that each sound does not lead to the resonance of only one narrow segment of the basilar membrane (a local Dirac function response) but initiates a traveling wave along the length of cochlea. Different frequencies of sound produce different traveling waves and these, in turn, have *peak* amplitudes at different points along the basilar membrane. Each frequency, therefore, produces a peak at a different position (exactly where Helmholtz had predicted), but, contrary to the resonance theory, there is an extension to the sides of the peak. This *representation* of the stimuli in visual [26] and other sensory systems [6] creates a clear advantage over "Dirac delta function" representation. There is only a finite number of receptors that can respond (see, hear, etc.) to a specific portion of external stimuli and the system is faced with the dilemma of a potential undersampling (too few receptors for a given area, e.g., - a visual patch). Should the "Dirac" representation be used by the visual system, it would imply that the localization of points in space will be limited by the resolution of the retina (number of sensors per area of visual field). On the other hand, if the representation is encoded by an aggregate (for the eye this corresponds to a group of cones or rods), the location of a stimulus may be resolved beyond the resolution of the receptive field associated with each individual sensory cell (cone or rod), as is illustrated in Figure 2.6. While the first encoding scheme introduces uncertainty associated with the width of the sensor, the second (specifically sensors labeled A and B), will allow to pinpoint the location of the stimuli **within** sensor X.

Using the intensity function similar to a Gaussian to reconstruct the coordinates of the spatial point results in the reconstruction of a spatial disc tangent to the distance line. The coordinates of all the points comprising that disc are erroneous except for a single central point of that disc for which the computations are physically correct. This situation represents a common perceptual experience. The reader can verify the

last statement by observing discrete light sources (stars or road lights) with blurred vision. They will appear to him as discs even though the light sources subtend incomparably small angles. Also note that the cut of a specific intensity function with the spherical retina is not position invariant but undergoes deformations which depend on eccentricity of the projection from the optical axis. This effect did not contaminate the present simulations because the latter were restricted to the kinematical aspect of the reconstruction process; it should, however, be taken into account when motion fields are extracted from a succession of real images. If one wants to use the present differential approach to solve for realistic situations where the scene is comprised of multiple spatial points, then one should consider two different points, R_1 and R_2, with objective distances ρ_1 and ρ_2 from the pinhole. Each of the two points projects a light intensity function. These would appear as two independent continuously differentiable intensities centered at the retinal coordinates θ_1, ϕ_1 and θ_2, ϕ_2 respectively. Now one should consider the superposition of these two independent signals as an input for the motion extraction mechanism.

The novelty of the present work stems mainly from the essential differences that exist between the current approach and the alternative solution given apparently for a single point [1] and advanced by Longuet-Higgins and Prazdny [10]. First, the critique raised by Bruss and Horn [1] against Longuet-Higgins and Prazdny's work holds because the latter work assumed that the optic flow always contains a straight-line (or great circle) trajectory. In visual reality this is not always the case because normally when we fixate there is a considerable amount of torsional eye movements [7]. Torsional component in the eye's angular velocity vector rules out the creation of such straight-line trajectories in the optic flow. Thus, this model applies only to a special case in which straight line in the flow field is detected. Second, the model of the latter authors assumed that the visual scene is a continuous textured surface. Even though the projection of each surface point is a Dirac delta function, the continuity of the visual surface assures that the retinal projection is continuous too and every retinal point has a small neighborhood. This assumption justifies taking the spatial derivative of the optic flow

but requires, in principle, a point and its near environment which eventually makes more than one point. To overcome these drawbacks, we first considered the scene as an ensemble of discrete point light sources each having its own distance ρ from the pinhole. Since the distance line of a discrete point has no spatial extension, then $\rho_\theta = \rho_\phi = 0$. When the scene was assumed to be comprised of continuous surface, we derived enough independent equations to eliminate the elements of the gradients of the surface (see *Appendix B*). Therefore, the projection of a single point regardless of its structure was sufficient to obtain a solution. In our case, taking the spatial derivatives of the flow equations was justified by the assumption about the intensity point spread function. The light intensity function creates a small neighborhood to each retinal projection even when it is emanating from a spatial point subtending infinitesimally small angle [35].

The relevance of the present derivations to human vision becomes more apparent when one considers the autokinetic phenomenon. This effect in general, and the latency interval in particular, shows that the capacity of the human visual system is far beyond what was suggested by the most rigorous computational models. Here we argue that passive navigation capacity as contrasted to active navigation capacity is essential to account for visual stability of a single point. First, we define active navigation as a 3-D reconstruction process that utilizes extraretinal signals in addition to retinal signals. Second, we note that in the case of active navigation, the indeterminate scale problem does not apply, in principle, to visual perception because there are other sources of information about the motion of the eye in space. Assuming the extraretinal signals as an additional source of information, then there are several reasons for why they cannot fully explain visual stability. First, even when the head is fixed and the eye is fixating a static point, there are involuntary eye movements that generate permanent slips of the retinal image [5]. Second, the level of physiological noise known to exist in the oculomotor signals [41] cannot perfectly compensate for the retinal slips of the image. Third, retinal slips due to head movements are only partially compensated because the vestibular system is insensitive to linear motion [15]. Taking all these reasons

together, one must conclude that a considerable amount of image instability should normally exist would visual stability being carried out solely on the basis of extraretinal information. Moreover, it has been shown that the predicted retinal slips due to noise in the oculomotor signals alone, are much larger than the spatial and temporal resolution limits of the visual system in detecting comparable slips [3, 22]. Thus, these slips, being not perfectly compensated, should be visible. Since they are not normally visible, the inevitable conclusion is that they are stabilized more accurately by the retinal signals via passive navigation computations. Indeed, this is what all the simulation have shown. Introducing measurement error to the observable quantities promptly generated jittery movements to the reconstructed point. A smooth unpredictable excursion typical to autokinetic movement was best simulated by introducing oculomotor noise.

Chapter 3

Six Degrees of Freedom: General Motion

3.1 The Retinal Flow Equations for Six Degrees of Freedom

We now extend the model such that any general motion of the observer is allowed. Since the pure rotations constitutes just a special type of general motion, many previous results are still applicable to the six degrees of freedom. Therefore, after the extended flow equations are derived, only differential (linear) case will be considered. As can be seen from the derivations provided in the *Appendix D*, the discrete approach gives a solution to the same unknowns and does not provide any additional information.

Using the exact same conventions as before, we start with any visual point as defining an esoteric inertial space with an origin not necessarily coinciding with the point itself. The major difference is that the observer's motion may consist of up to 6 independent degrees of freedom denoted by $\mathbf{v}$ and ω; where $\mathbf{v} = [v_x(t), v_y(t), v_z(t)]$ is the translatory velocity of the eye coordinate system, and $\omega = [\omega_x(t), \omega_y(t), \omega_z(t)]$ is the angular velocity of the eyeball. Since the eye coordinate system no longer coincides with the world coordinate system, the new situation may now be represented as shown in Figure 3.1. Thus, we have:

$$\mathbf{R} = \mathbf{r} + \mathbf{R}_0, \tag{3.1}$$

where $\mathbf{R}$ is the position vector of object point in the point space, $\mathbf{r}$ is the distance between the eye's system origin and the point, and $\mathbf{R}_0$ is a vector connecting the point's system origin with the eye's system origin. Then, the visual point $\mathbf{R}$ can be given in terms of the eye system notation $\mathbf{r}$ by the transformation:

$$\mathbf{r} = -\mathbf{R}_0 + \mathbf{M}\mathbf{R}, \tag{3.2}$$

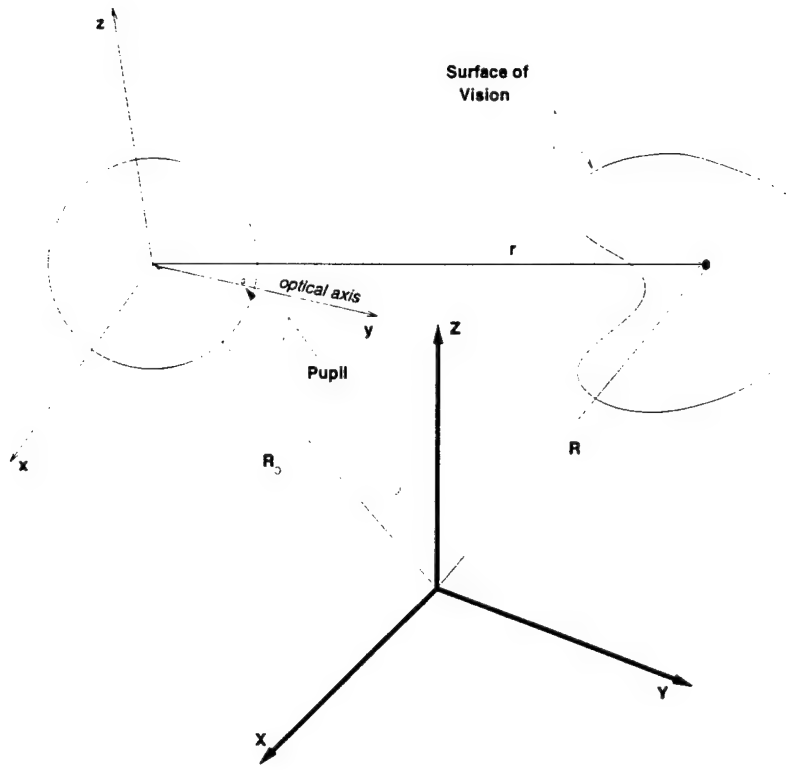


Figure 3.1: Definition of systems of coordinates in the SPIN theory for the general motion (rotation and translation) case. Note that the origins of X, Y, Z and x, y, z coordinate systems no longer coincide.

where $\mathbf{M}$ is an orthogonal transformation matrix the elements of which are λ, μ, and ν as before. The absolute time derivative of $\mathbf{R}$ with respect to time is then:

$$\frac{d\mathbf{R}}{dt} = \frac{d\mathbf{R}_0}{dt} + \frac{d\mathbf{r}}{dt} + \omega \times (\mathbf{R}_0 + \mathbf{r}) = 0, \tag{3.3}$$

where the right hand of the equation is given in the inertial coordinate system, while the right hand side is in the eye coordinate system. Since $\frac{d\mathbf{R}_0}{dt} + \omega \times \mathbf{R}_0$ is the velocity of the eye, $\mathbf{v}$, in the eye system, we then have:

$$\frac{d\mathbf{R}_0}{dt} + \omega \times \mathbf{R}_0 + \frac{d\mathbf{r}}{dt} + \omega \times \mathbf{r} = \mathbf{v} + \frac{d\mathbf{r}}{dt} + \omega \times \mathbf{r} = 0. \tag{3.4}$$

Expanding equation (3.4) and substituting for $\dot{x}$, $\dot{y}$ leads to a the following equations:

$$\begin{aligned} v_x + \dot{\rho} sin\frac{\theta}{2} sin\phi + \rho\frac{\dot{\theta}}{2} cos\frac{\theta}{2} sin\phi + \rho\dot{\phi} sin\frac{\theta}{2} cos\phi + \\ \omega_y \rho sin\frac{\theta}{2} cos\phi - \omega_z (c + \rho cos\frac{\theta}{2}) &= 0 \end{aligned} \tag{3.5}$$

$$\begin{aligned} v_y + \dot{\rho} cos\frac{\theta}{2} - \rho\frac{\dot{\theta}}{2} sin\frac{\theta}{2} + \omega_z \rho sin\frac{\theta}{2} sin\phi - \\ \omega_x \rho sin\frac{\theta}{2} cos\phi &= 0 \end{aligned} \tag{3.6}$$

$$\begin{aligned} v_z + \dot{\rho} sin\frac{\theta}{2} cos\phi + \rho\frac{\dot{\theta}}{2} cos\frac{\theta}{2} cos\phi - \rho\dot{\phi} sin\frac{\theta}{2} sin\phi + \\ \omega_x (c + \rho cos\frac{\theta}{2}) - \omega_y \rho sin\frac{\theta}{2} sin\phi &= 0. \end{aligned} \tag{3.7}$$

Eliminating $\dot{\rho}$ in (3.5)-(3.7) and solving for $\dot{\theta}$ and $\dot{\phi}$, the basic flow equations are then:

$$\dot{\theta} = \frac{2}{\rho}(v_y sin\frac{\theta}{2} - (\rho + c cos\frac{\theta}{2})(\omega_x cos\phi - \omega_z sin\phi) - cos\frac{\theta}{2}(v_x sin\phi + v_z cos\phi)), \tag{3.8}$$

$$\dot{\phi} = (c + \rho cos\frac{\theta}{2})\frac{\omega_z cos\phi + \omega_x sin\phi}{\rho sin\frac{\theta}{2}} + \frac{v_z cos\phi - v_x sin\phi}{\rho sin\frac{\theta}{2}} - \omega_y. \tag{3.9}$$

Reader may easily verify that should $v_x = v_y = v_z = 0$, the above flow equations will become exactly the same as for the pure rotations (equations (2.13) and (2.14)).

3.2 Solution of the Flow Equations

Equations (3.8) and (3.9) can now be differentiated twice with respect to θ and ϕ, producing additional 12 equations that are listed in the *Appendix C* (assuming $\rho_\theta =$

$\rho_\phi = 0$). After eliminating dependent equations, one can solve for the angular velocity of the eye, giving:

$$\omega_x = Asin\phi - Bcos\phi, \tag{3.10}$$

$$\omega_y = -\dot{\phi} - \dot{\phi}_{\phi\phi}, \tag{3.11}$$

$$\omega_z = Bsin\phi + Acos\phi, \tag{3.12}$$

$$\tag{3.13}$$

where

$$A = \frac{cos\frac{\theta}{2}}{sin\frac{\theta}{2}}\dot{\phi}_{\phi\phi} - 2\dot{\phi}_\theta, \tag{3.14}$$

$$B = 2\dot{\theta}_{\theta\theta} + \frac{\dot{\theta}}{2}. \tag{3.15}$$

Also, the following visual field *invariances* associated with the distance to the visual point and the motion of the eye in space can be computed:

$$\frac{v_a}{\rho} = \frac{-\dot{\theta}_{\theta\phi}}{sin\frac{\theta}{2}} \tag{3.16}$$

$$\frac{v_b}{\rho} = -Bcos\frac{\theta}{2} - \dot{\phi}_\phi sin\frac{\theta}{2} \tag{3.17}$$

$$\frac{v_y}{\rho} = \frac{\dot{\theta}_{\phi\phi} + \dot{\theta}}{2sin\frac{\theta}{2}} \tag{3.18}$$

where

$$v_a = cA + E, \tag{3.19}$$

$$v_b = cB - D, \tag{3.20}$$

$$D = v_z cos\phi + v_x sin\phi, \tag{3.21}$$

$$E = v_z sin\phi - v_x cos\phi. \tag{3.22}$$

Since the distance ρ is not known, the computational model can not solve for v_a, v_b, or v_y directly. This implies that there is no hope to obtain components D or E, which are necessary to solve for

$$v_x = Dsin\phi - Ecos\phi, \tag{3.23}$$

$$v_z = Esin\phi + Dcos\phi. \tag{3.24}$$

The three invariances in (3.16) - (3.18) clearly are not enough to solve for the four unknowns: ρ, v_x, v_y, and v_z. Therefore, some additional constraints must be imposed on the motion, which we will investigate next.

3.3 Biologically Plausible Constraints

The search for additional constraints can be substantially reduced if we consider the global characteristics of the problem at hand. Since it is possible to compute the rotational components, **omega**, from the optic flow for **any** general motion, we should look at the effects of the translation and rotation separately. A common problem that plagued other solutions is that for any given linear motion, say $\mathbf{V}$, and any distance to some visible point, say $\mathbf{r}$, we can multiply them both by some constant, say Q, such that the new motion is now $Q\mathbf{V}$ and the distance is $Q\mathbf{r}$. As a result, in the previous models the observed effect on the optic flow generated by the former was indistinguishable from the latter, a problem known as *indeterminate scale problem.*

In the SPIN theory one distinguishing element is a fixed radius of the eyeball, c. While magnitudes of the motion may be changed, the radius remains fixed. Therefore, consider the following situation (Figure 3.2). Let both, - the rotation and the translation, be along the same line, - say z axis. Then, the slip of the image will be a function of two components: the drift along the z axis due to translation and, at the same time, a drift along the x axis due to rotation. Here, doubling the distance to the visible point and at the same time doubling the translation vector will result in the same drift along the z axis, but the drift along the x axis (due to rotation), will remain the same as long as the radius of the eyeball remains constant in both cases. The only two possible ways to produce exactly same displacement on the retina are:

a) to compensate with an appropriate translation along the y axis (forward or backward). (Note that in this case the final position of the eyeball will coincide with the previous case.)

b) to compensate with an appropriate translation along the x axis (left or right).

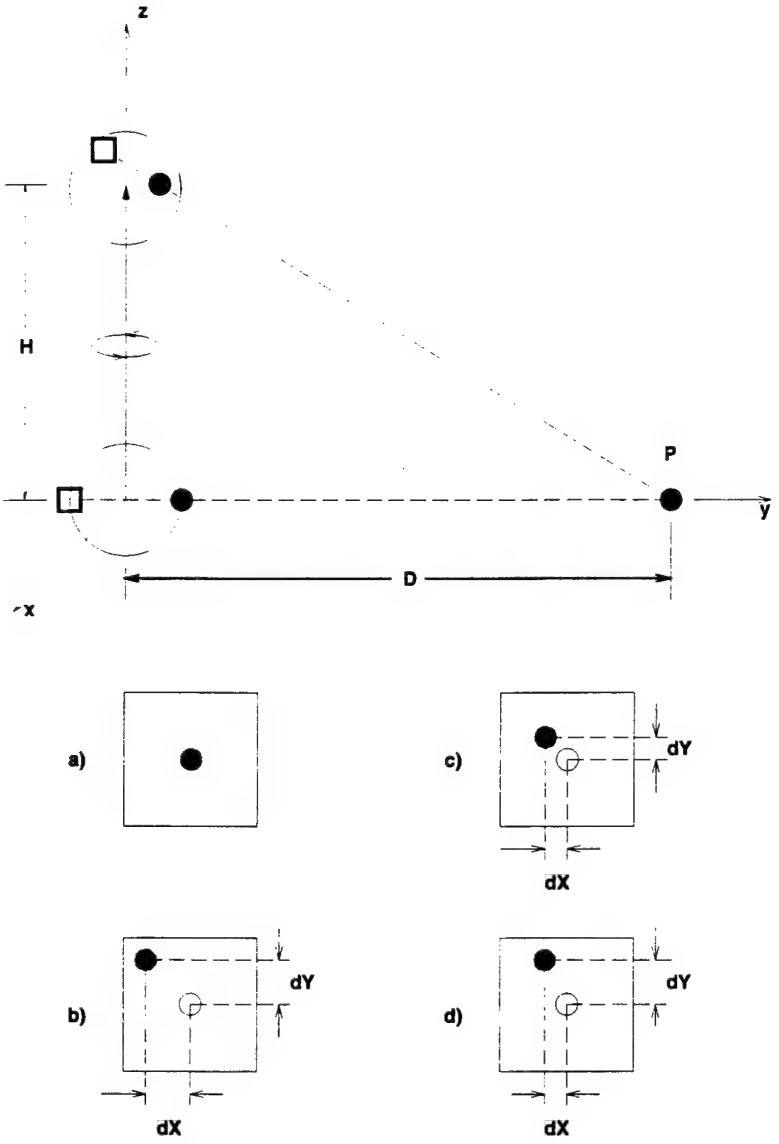


Figure 3.2: Projection of a point (P) onto a small segment of a retina: a) initial projection; b) after vertical motion and rotation (shaded); c) the same motion but distance (D) is doubled; d) after distance (D) and linear motion doubled, while rotational component is left unchanged. See text for details.

In either case above, what is required is a change in the **direction** of the translation vector. This directional sensitivity of the system, therefore, identify the two natural constraints that may be used to solve the problem: 1) explore some well defined relation between direction of the rotation vector and the translation vector, or, 2) explore some well defined property of a direction associated with the translation vector alone.

The first suggestion points to vestibulo-ocular reflex (VOR). Traditionally, it has been described as a distinct, phylogenetically old oculomotor subsystem, which serves to stabilize the gaze direction. It is suppose to act as a stereotyped reflex with definite input-output relations. As a reflex, it should also show when a subject is rotated passively in the dark. On the basis of eye movement measurements in the dark, however, Collewijn [11] argued that the traditional view of the VOR is not realistic. He suggests a more fruitful hypothesis that the vestibular signals are just one of many inputs to a spatial localization process, which computes the relative position (and motion) between the subject and the target of his choice. If, in addition, VOR provide the angle between the rotational and translational components of the motion (by keeping it fixed), say ***alpha***, then there is an additional relationship that can be used by the system

$$\mathbf{V} \cdot \omega = |\mathbf{V}||\omega| cos\alpha. \tag{3.25}$$

This leads to a second order equation in ρ (see *Appendix E*). It turns out that out of the two possible roots, one always places the visual point **within** the radius of the eyeball, while the other gives the true distance to the point. Thus, it enables the system to resolve the ambiguity by choosing the solution that is larger than **c** (radius of the eye).

Unfortunately, one must be very careful in exploiting this constraint. First, VOR involves choosing some visible point in the scene (fixation point), and minimizing retinal slips of this point on the retina ($\dot{\theta} = \dot{\phi} = 0$). This is clearly an active task and falls outside of the domain of passive navigation. Second, if the angle that VOR is trying to preserve is 90 degrees, then $cos(\alpha) = cos(90) = 0$ and equation (3.25) above becomes homogeneous. As such, it does not provide any additional information and, therefore, we are forced to avoid the orthogonality between v and ω.

The other possible constraint stems from the fact that the mass of the observer's head is much larger than the mass of the eye. Therefore, it is expected that for a sequence of views, while the rotation vector varies, the direction of the translational velocity in the world coordinate system remains approximately fixed (quasi-linear). Since the solution for the rotational component allows us to compute the exact orientation of the observer at any moment in time, it becomes a simple task to predict the direction of the translation vector for a sequence of frames when its direction remains approximately fixed. Formally, we than have the following new equations:

$$\frac{d\omega}{dt} = f(t) \neq 0 \tag{3.26}$$

while

$$\frac{d\mathbf{V}}{dt} = 0 \tag{3.27}$$

This produces a system that can be solved either by differentiating the flow equations with respect to time or, alternatively, using three (or more) views. The latter is similar to the discrete case, except instead of using multiple points we are using multiple views of the same point. For a linear constant motion, then, it requires three views, for the linear acceleration, - four views, etc. (see *Appendix F*) for the example of linear motion solution).

3.4 Discussion

So far, we have dealt with the need of passive navigation capacity to account for visual stability. The mathematical derivations have shown that the problem of visual stability cannot be separated from the problem of absolute distance perception because the amount of retinal motion depends also on the distance of the point from the pinhole. Both problems are reconciled by the SPIN theory with the assumption about the intrinsic metric scale. The metric scale enabled accurate decomposition of changes in retinal projection into their distance and eye rotation components. The significance of metric scale in mental representation is discussed next.

Everyday experience in perceiving constant object shapes from changing perspectives in varying contexts indicates that vision acquires information about the spatial structure of environmental objects with impressive speed, reliability, precision, and stability. Geometrical optics, however, predicts that the retinal image of objects are changing and deforming from moment to moment. One interpretation of the perceived constancy of position and size is given by Gibson [8, 9] and is based on the conjecture that dynamics retinal projections preserve certain invariants, and higher order variables which are picked up and yield a direct perception of the object's constancies. While the Gibson's explanation is of a qualitative nature, the modern computational models facing the indeterminate-scale problem have tried to solve the perceptual constancy phenomenon with affine geometry [39, 25]. But, the reconstruction of 3D objects in these models is restricted to an arbitrary affine transformation, and therefore yields only relative depth perception. This type of solution does not fully account for the perceptual constancy of position and size and raises the question of performance without a veridical scale.

The question of metric scale in mental representation may be ignored or disregarded in several ways: One may argue that physical scales with regard to the mind's eye are meaningless because the mind is a metaphysical entity and physical scales are irrelevant. Or alternatively, to argue that the mind's eye is tolerant with certain distance and magnification paradoxes typical to prosthetic vision situations [13, 19]. These situations evoke image instability that may result from modified metric scale [12]. The plausibility of this solipsistic view evaporates when one considers motion sickness that results when the visual input contradicts a veridical kinesthetic input. One may argue that space perception is not always veridical (or never veridical) by referring to the endless list of known visual illusions. Also one may assume that the perceived space is represented at the mind's eye in an arbitrary scale. Then each observer may have her esoteric scale. These types of solutions immediately raise epistemological questions about visiomotor coordination within the organism as well as how successful communication can be carried out between different organisms having different scales (e.g., when two people have

to reach, manipulate and/or carry an extended object). Another theoretical solution is to scale the perceived world on the basis of external information. As for the human visual system, it was postulated that motor interaction with the physical objects may calibrate and converge the representation to veridical metric scale (this postulate was raised by the empiricistic school as a supplement to the classical cue theory of space perception). However, no one has shown mathematically how the kinesthetic sense is scaled. Thus this explanation remains circular because the scaling problem applies to kinesthetic sense as well as to the visual sense. Finally, one can suggest that the representation of objects may be calibrated on the basis of other external magnitudes such as the height of the eyes above the ground [24]. Note that this type of explanation, due to the indeterminacy of the solution, cannot tell whether the subjective scales of different observers are the same.

The SPIN theory suggests a simple way to break this circularity by the assumption about the existence of intrinsic metric scales. In monocular vision the radius of the eyeball serves as one of these scales. Recent measurements on the ocular occlusion phenomenon have shown that in human observers this magnitude is about 11mm30. Furthermore, when an eye in 6 degrees of freedom is considered, the distance between the origins of the eye and head systems presumably serves as an additional scale [12]. The latter distance is much larger magnitude if one centers the origin of the head-system in the ipsilateral vestibular organ. Analogously, for binocular vision the distance between the eyes, known as the inter-ocular distance, is the scale. A psychophysical support for the latter conjecture was recently obtained by showing that the perceived depth in random dot stereogram was highly correlated with the observer's inter-ocular distance [13, 18]. Furthermore, depth judgments of 45 subjects fall nicely on the theoretical curve calculated from a formula that is based on triangulation. This finding has received additional support by neurophysiological measurements showing that cells in V1 of monkeys respond to the absolute depth of the display [40]. While it is simpler to show that the inter-ocular distance may have a role in establishing metric scale in stereopsis, it is a much more difficult task to show that same applies to monocular vision.

Currently, we plan to pursue analogous measurements with monocular vision. However, given that the dependence of stereoscopic depth on the inter-ocular-distance indicates the existence of metric scale in binocular vision, it is unlikely that the monocular mechanism utilizes an indeterminate-scale solution, as suggested by the common view. Thus, we conclude that the visual system as a whole reconstructs objects with a metric scale. The present work shows that this type of solution is mathematically feasible. Moreover, the assumption made by the SPIN theory about the metric representation nicely solves the position and size constancy problem because if the mental representation is given in terms of object's coordinates undergoing linear transformation, then the position and the distance function are preserved.

Appendix A

Alternative Basis

Generally, the motion parameters of an object through space can be conveniently specified by a two three-dimensional vectors: $\omega = [\omega_x, \omega_y, \omega_z]$ and $\mathrm{v} = [v_x, v_y, v_z]$. Due to the model that was chosen to mimic the mechanics of the eye (a pinhole camera), it is more convenient to use an alternative representations that are specific to the projection of a specific point onto the retina. The alternative velocity parameters of an observer associated with any projection $p = [\theta, \phi]$ form then an alternative basis:

$$\omega_p = \begin{bmatrix} A \\ \omega_y \\ B \end{bmatrix} \tag{A.1}$$

$$\mathbf{V}_p = \begin{bmatrix} D \\ \omega_y \\ E \end{bmatrix}, \tag{A.2}$$

where

$$A = \omega_x sin\phi + \omega_z cos\phi, \tag{A.3}$$

$$B = \omega_z sin\phi - \omega_x cos\phi, \tag{A.4}$$

$$D = v_z cos\phi + v_x sin\phi, \tag{A.5}$$

$$E = v_z sin\phi - v_x cos\phi. \tag{A.6}$$

It is very simple to verify that the basis transformation can be directly obtained by:

$$M_b \cdot \omega_p = \omega \tag{A.7}$$

and

$$M_b \cdot \mathbf{v_p} = \mathbf{v} \tag{A.8}$$

where

$$M_b = \begin{bmatrix} sin\phi & 0 & -cos\phi \\ 0 & 1 & 0 \\ cos\phi & 0 & sin\phi \end{bmatrix} \tag{A.9}$$

The inverse transformation is then given by using the inverse of the matrix M_b, which is, since M_b is orthogonal, just a its transpose (${M_b}^{-1} = {M_b}^T$). Using ω_p and $\mathbf{v}_p$ simplifies the mathematical expressions and, since they form the **true basis** (solving for one provides the solution for the other), they will be used in the following derivations.

Appendix B

A Solution for a Single Point on a Continuous Surface

We assume that the objects of vision are composed of continuous surfaces. Differentiating equations (2.13) and (2.14) with respect to θ and ϕ yields:

$$\dot{\theta}_\theta = -\frac{c}{\rho}Bsin\frac{\theta}{2} - 2cB\frac{\rho_\theta}{\rho^2}cos\frac{\theta}{2}, \tag{B.1}$$

$$\dot{\theta}_\phi = 2A\left(1+\frac{c}{\rho}cos\frac{\theta}{2}\right) - 2cB\frac{\rho_\theta}{\rho^2}cos\frac{\theta}{2}, \tag{B.2}$$

$$\dot{\phi}_\theta = -\frac{A}{2sin^2\frac{\theta}{2}}\left(1+\frac{c}{\rho}cos\frac{\theta}{2}+2c\frac{\rho_\theta}{\rho^2}sin\frac{\theta}{2}\right), \tag{B.3}$$

$$\dot{\phi}_\phi = -\frac{B}{sin\frac{\theta}{2}}\left(cos\frac{\theta}{2}+\frac{c}{\rho}\right) - \frac{cA\rho_\phi}{\rho^2 sin\frac{\theta}{2}}. \tag{B.4}$$

The gradient components ρ_θ and ρ_ϕ can be eliminated in (B.1) - (B.4), yielding the following two independent equations:

$$A\dot{\theta}_\theta - B\dot{\phi}_\theta sin\theta = \frac{AB}{sin\frac{\theta}{2}}\left(cos\frac{\theta}{2}+\frac{c}{\rho}cos\theta\right), \tag{B.5}$$

$$A\dot{\theta}_\phi - B\dot{\phi}_\phi sin\theta = 2\frac{c}{\rho}cos\frac{\theta}{2}\left(A^2+B^2\right)+2A^2+2B^2cos^2\frac{\theta}{2}. \tag{B.6}$$

Equations (2.13), (2.14), (B.5), and (B.6) constitute a set of four independent nonlinear equations with four unknowns that is valid for a single point irrespective of surface structure. The solution for this set of equations is obtained in the following way: first, expressions A, B, A^2, and B^2 are substituted into equations (B.5) and (B.6). Then an expression for $\dot{\phi}+\omega_y$ is isolated from equation (2.14) and substituted into equation (B.5), leading to the following expression in ρ:

$$\dot{\theta}_\phi sin\tfrac{\theta}{2} - \frac{\dot{\theta}_\theta\dot{\phi}_\phi}{\dot{\phi}_\theta}sin\tfrac{\theta}{2} + \frac{\dot{\theta}\dot{\phi}_\phi}{2\dot{\phi}_\theta}cos\tfrac{\theta}{2} - \frac{\dot{\theta}\dot{\theta}_\theta}{2\dot{\phi}_\theta}cos\tfrac{\theta}{2} + \frac{\dot{\theta}^2cos^2\frac{\theta}{2}}{4\dot{\phi}_\theta sin\frac{\theta}{2}} = \frac{\dot{\theta}c\left(\dot{\theta}_\theta+\dot{\phi}_\phi\right)}{2\dot{\phi}_\theta\left(\rho+ccos\frac{\theta}{2}\right)} + \frac{\dot{\theta}^2c^2sin^3\frac{\theta}{2}}{4\dot{\phi}_\theta\left(\rho+ccos\frac{\theta}{2}\right)^2} + \frac{2\dot{\theta}\dot{\phi}_\theta}{\left(2\dot{\theta}_\theta-\dot{\theta}cot\frac{\theta}{2}+\frac{\dot{\theta}csin\frac{\theta}{2}}{\rho+ccos\frac{\theta}{2}}\right)sin\theta sin\frac{\theta}{2}}. \tag{B.7}$$

The left-hand side of (B.7) comprises only observable quantities and the compound unknown $\rho + ccos\frac{\theta}{2}$. Substituting $\psi = \rho + ccos\frac{\theta}{2}$ into equation (B.7) produces a cubic polynomial in ψ, where all the coefficients are combinations of observable quantities. Solving this polynomial for ψ yields three roots: one is real and the other two are imaginary conjugates. We take the real root as a solution for calculating ρ. The other unknowns are then solved for using equations (2.13) and (2.14), and either (B.5) or (B.6).

Even though this solution takes into account all aspects of retinal projection for distance calculations, it is unlikely that a biological system can utilize it. The solution is, however, applicable to machine vision.

Appendix C

Flow Equations for the General Motion

Assuming $\rho_\phi = \rho_\theta = 0$, differentiation of equations (3.8) and (3.9) twice with respect to θ and ϕ yield the following additional twelve flow equations:

$$\dot{\theta}_\theta = \frac{1}{\rho}\left[-sin\frac{\theta}{2}(cB - D) + v_y cos\frac{\theta}{2}\right], \tag{C.1}$$

$$\dot{\theta}_{\theta\phi} = -sin\frac{\theta}{2}\frac{cA + E}{\rho}, \tag{C.2}$$

$$\dot{\theta}_{\theta\theta} = \frac{1}{2\rho}\left[-cos\frac{\theta}{2}(cB - D) - v_y sin\frac{\theta}{2}\right], \tag{C.3}$$

$$\dot{\theta}_\phi = \frac{2}{\rho}\left[\left(\rho + ccos\frac{\theta}{2}\right)A + cos\frac{\theta}{2}E\right], \tag{C.4}$$

$$\dot{\theta}_{\phi\theta} = -sin\frac{\theta}{2}\frac{cA + E}{\rho}, \tag{C.5}$$

$$\dot{\theta}_{\phi\phi} = \frac{2}{\rho}\left[cos\frac{\theta}{2}D - \left(\rho + ccos\frac{\theta}{2}\right)B\right], \tag{C.6}$$

$$\dot{\phi}_\theta = \frac{\left(\rho + ccos\frac{\theta}{2}\right)A + cos\frac{\theta}{2}E}{-2\rho sin^2\frac{\theta}{2}}, \tag{C.7}$$

$$\dot{\phi}_{\theta\phi} = \frac{cos\frac{\theta}{2}D - \left(\rho + ccos\frac{\theta}{2}\right)B}{-2\rho sin^2\frac{\theta}{2}}, \tag{C.8}$$

$$\dot{\phi}_{\theta\theta} = \frac{2Acos\frac{\theta}{2}\rho + (cA + E)\left(1 + cos^2\frac{\theta}{2}\right)}{4\rho sin^3\frac{\theta}{2}}, \tag{C.9}$$

$$\dot{\phi}_\phi = \frac{D - \left(c + \rho cos\frac{\theta}{2}\right)B}{\rho sin\frac{\theta}{2}}, \tag{C.10}$$

$$\dot{\phi}_{\phi\theta} = \frac{\left(\rho + ccos\frac{\theta}{2}\right)B - cos\frac{\theta}{2}D}{2\rho sin^2\frac{\theta}{2}}, \tag{C.11}$$

$$\dot{\phi}_{\phi\phi} = -\frac{\left(c + \rho cos\frac{\theta}{2}\right)A + E}{\rho sin\frac{\theta}{2}}. \tag{C.12}$$

Equations (3.8), (3.9), and (C.1) - (C.12) are independent in A, B, ω_y, $\frac{(cB-D)}{\rho}$,

$\frac{(cA+E)}{\rho}$, and $\frac{v_y}{\rho}$. Therefore, multiplying by matrix M_p, we can solve directly for ω_x, ω_z, ω_y, $\frac{c\omega_x+v_z}{\rho}$, $\frac{c\omega_z-v_x}{\rho}$, and $\frac{v_y}{\rho}$. Since $\frac{(cB-D)}{\rho}$, $\frac{v_y}{\rho}$, and $\frac{(cA+E)}{\rho}$ are **inseparable**, the flow equations can not be solved for either D, E, v_y, or ρ.

Appendix D

General Motion, Discrete Case Along Same ϕ

For mathematical simplicity, let us use points along same ϕ. Also, we need the following substitutions:

$$v_a = cA + E, \tag{D.1}$$

and

$$v_b = cB - D. \tag{D.2}$$

Solving equations (3.8) and (3.9) for ρ and equating the two, we have:

$$v_y\left(\dot{\phi}+\omega_y\right)tan\frac{\theta}{2} - v_a\frac{\dot{\theta}-2B}{sin\theta} + v_b\left(-Acot\frac{\theta}{2}+\dot{\phi}\right) = v_yA - (cB-D)\omega_y. \tag{D.3}$$

The following definitions simplify the above expression:

$$G_1 = \frac{v_y}{v_a}, \tag{D.4}$$

$$G_2 = \frac{v_y\omega_y}{v_a}, \tag{D.5}$$

$$G_3 = \frac{v_bA}{v_a}, \tag{D.6}$$

$$G_4 = \frac{v_b}{v_a}, \tag{D.7}$$

$$G_5 = \frac{v_yA-(cB-D)\omega_y}{v_a}. \tag{D.8}$$

Thus, after rearranging terms, equation (D.3) can be written as:

$$\frac{2}{sin\theta}B + tan\frac{\theta}{2}G_1\dot{\phi} + tan\frac{\theta}{2}G_2 - \frac{\dot{\theta}}{sin\theta} - G_3cot\frac{\theta}{2} + G_4\dot{\phi} = G_5. \tag{D.9}$$

Taking two different points (along same ϕ), say i and j, results in two equations in the form of (D.9). Subtracting one from the other yields:

$$\left(\frac{2}{sin\theta_i} - \frac{2}{sin\theta_j}\right)B + \left(tan\frac{\theta_i}{2}\dot{\phi}_i - tan\frac{\theta_j}{2}\dot{\phi}_2\right)G_1 +$$

$$\left(tan\frac{\theta_i}{2} - tan\frac{\theta_j}{2}\right) G_2 + \left(-cot\frac{\theta_i}{2} + cot\frac{\theta_j}{2}\right) G_3 + \left(\dot{\phi}_i - \dot{\phi}_2\right) G_4 - \frac{\dot{\theta}_i}{sin\theta_i} + \frac{\dot{\theta}_2}{sin\theta_j} = 0. \quad \text{(D.10)}$$

where subscripts i and j refers to a specific point. Thus, for any two general point, say i and j, define the following functions:

$$F_{01}(i,j) = \frac{2}{sin\theta_i} - \frac{2}{sin\theta_j}, \quad \text{(D.11)}$$

$$F_{02}(i,j) = tan\frac{\theta_i}{2}\dot{\phi}_i - tan\frac{\theta_j}{2}\dot{\phi}_j, \quad \text{(D.12)}$$

$$F_{03}(i,j) = tan\frac{\theta_i}{2} - tan\frac{\theta_j}{2}, \quad \text{(D.13)}$$

$$F_{04}(i,j) = -cot\frac{\theta_i}{2} + cot\frac{\theta_j}{2}, \quad \text{(D.14)}$$

$$F_{05}(i,j) = \dot{\phi}_i - \dot{\phi}_j, \quad \text{(D.15)}$$

$$F_{06}(i,j) = -\frac{\dot{\theta}_i}{sin\theta_i} + \frac{\dot{\theta}_j}{sin\theta_j}. \quad \text{(D.16)}$$

Now, overwriting (D.10) for any two point i and j, yields:

$$F_{01}(i,j)B + F_{02}(i,j)G_1 + F_{03}(i,j)G_2 + F_{04}(i,j)G_3 + F_{05}(i,j)G_4 + F_{06}(i,j) = 0. \quad \text{(D.17)}$$

Equation (D.17) can now be solved for B. Using three points, say i, j, and k, one can get two independent equations of the form of (D.17) (e.g., - one for i and j, and the other for i and k). Subtracting one from the other (effectively solving for $B - B$), results in:

$$\left(\frac{-F_{02}(i,j)}{F_{01}(i,j)} + \frac{F_{02}(i,k)}{F_{01}(i,k)}\right) G_1 + \left(\frac{-F_{03}(i,j)}{F_{01}(i,j)} + \frac{F_{03}(i,k)}{F_{01}(i,k)}\right) G_2 + \vdots \quad \text{(D.18)}$$

$$\frac{-F_{06}(i,j)}{F_{01}(i,j)} + \frac{F_{06}(i,k)}{F_{01}(i,k)} = 0.$$

The pattern becomes now evident. We rename terms for G_1 through G_4 and the constant term at the end to be some new functions, say $F_11(i,j,k)$ through $F_15(i,j,k)$. Solve for G_1, add a new point, and then subtract two equations. Continuing in this fashion, produces **exactly** the same results as in differential case. A minimum of five points are necessary to solve for A, B, ω_y, $\frac{v_a}{v_b}$, and $\frac{v_y}{v_a}$. Any additional points beyond the initial five are redundant. This restriction manifests itself through the fact that some

G_i terms can not be solved for (the $F_{pq}(\cdots)$ coefficient become identically zero, as, for example, is in the case of G_3 above). Since the same expressions can be extracted in both, - discrete and differential cases, - it is only logical to conclude that the additional constraints described before are necessary to solve for the distance to the five points and the translational component of the observer's velocity.

Appendix E

Polynomial Solution for General Motion

Suppose that the angle between the vector of rotation (**omega**) and the translation vector (v) is somehow available to the system. Then, here we show how this additional constraint can be used to solve all remaining unknowns. To start, observe that the dot product of two vectors is given by:

$$\omega \cdot \mathbf{v} = \omega_x v_x + \omega_y v_y + \omega_z v_z = |\omega||\mathbf{v}|cos\alpha, \tag{E.1}$$

where α is the angle between the two vectors. Since the components of the translation vector can be expressed in terms of ρ, equation (E.1) can be rewritten in the form:

$$C_a\rho^2 + C_b\rho + C_c = 0, \tag{E.2}$$

where

$$\begin{aligned} C_a &= \left(\frac{cA+E}{\rho}\right)^2 + \left(\frac{cB-D}{\rho}\right)^2 + \left(\frac{v_y}{\rho}\right)^2 - \\ &\quad \left[\frac{-A(cB-D)+B(cA+E)+\omega_y v_y}{\rho}\right]^2 \frac{1}{|\omega|cos^2\alpha}, \qquad \text{(E.3)} \\ C_b &= -2cB\frac{(cB-D)}{\rho} - 2cA\frac{(cA+E)}{\rho}, \qquad \text{(E.4)} \\ C_c &= c^2B^2 + c^2A^2. \qquad \text{(E.5)} \end{aligned}$$

The coefficients C_a, C_b, and C_c can be readily computed. Therefore, ρ can be obtained by solving the polynomial equation (E.2). Although it is a second order polynomial, one of the roots always places the point inside or behind the eye. This breaks the ambiguity associated with this mathematical construct and the true distance (ρ) can be identified. By using this ρ, all other unknowns can be obtained from the other previously computed field invariances: $\frac{(cA+E)}{\rho}$, $\frac{(cB-E)}{\rho}$, and $\frac{v_y}{\rho}$.

Appendix F

Three Views Solution For General Motion

By using the basis transformation matrix M_b and field invariances $\frac{v_a}{\rho}$, $\frac{v_y}{\rho}$, and $\frac{v_b}{\rho}$, the new relations are derived:

$$\frac{c\omega_x + v_z}{c\omega_z - v_x} = \Lambda, \tag{F.1}$$

$$\frac{v_y}{c\omega_z - v_x} = \Gamma. \tag{F.2}$$

where Λ and Γ are computed from equations (3.16)-(3.18).

Rearranging terms, (F.1) and (F.2) can be rewritten as:

$$\Lambda v_x + v_z = N_1 \tag{F.3}$$

$$\Gamma v_x + v_y = N_2 \tag{F.4}$$

where N_1 and N_2 are $\Lambda c\omega_z - c\omega_x$ and $\Gamma c\omega_z$ respectively. Since the rotational component of the motion can be obtained from equations (3.10)-(3.15), the transformation matrix $\mathbf{M}$ can be computed at every step. Therefore, equations (F.3) and (F.4) can be expressed in the inertial space as:

$$\left(\Lambda^t m^t_{1,1} + m^t_{3,1}\right) V_x + \left(\Lambda^t m^t_{1,2} + m^t_{3,2}\right) V_y + \left(\Lambda^t m^t_{1,3} + m^t_{3,3}\right) V_z = N^t_1, \tag{F.5}$$

$$\left(\Gamma^t m^t_{1,1} + m^t_{2,1}\right) V_x + \left(\Gamma^t m^t_{1,2} + m^t_{2,2}\right) V_y + \left(\Gamma^t m^t_{1,3} + m^t_{2,3}\right) V_z = N^t_2. \tag{F.6}$$

where t represents time and $m^t_{i,j}$ are the elements of the $\mathbf{M}(t)$ rotation matrix. If the translational component, $\mathbf{V}$, is assumed to change linearly from view to view, then, by taking three views, we have four equations and four unknowns: tree components of $\mathbf{V}$ and some scalar linear acceleration component. This assumption enables the system to

solve for the scaled translation and the acceleration:

$$V_x = \frac{Q_{3,2}Q_{2,3}N_{1,1} - Q_{3,2}Q_{1,3}N_{2,1} - Q_{2,2}Q_{3,3}N_{1,1} + Q_{1,2}Q_{3,3}N_{2,1}}{DQ}, \quad \text{(F.7)}$$

$$V_y = -\frac{Q_{3,1}Q_{2,3}N_{1,1} - Q_{3,1}Q_{1,3}N_{2,1} - Q_{3,3}Q_{2,1}N_{1,1} + Q_{3,3}Q_{1,1}N_{2,1}}{DQ}, \quad \text{(F.8)}$$

$$V_z = \frac{Q_{3,1}Q_{2,2}N_{1,1} - Q_{3,2}Q_{2,1}N_{1,1} - Q_{3,1}Q_{1,2}N_{2,1} + Q_{3,2}Q_{1,1}N_{2,1}}{DQ}, \quad \text{(F.9)}$$

where

$$DQ = Q_{3,2}Q_{2,3}Q_{1,1} - Q_{3,2}Q_{1,3}Q_{2,1} - Q_{2,3}Q_{1,2}Q_{3,1} - Q_{2,2}Q_{3,3}Q_{1,1} + Q_{1,2}Q_{3,3}Q_{2,1} + Q_{1,3}Q_{2,2}Q_{3,1}, \quad \text{(F.10)}$$

$$Q_{1,i} = \Lambda_t m_{1,i}^t + m_{3,i}^t, \quad \text{(F.11)}$$

$$Q_{2,i} = \Gamma_t m_{1,i}^t + m_{2,i}^t, \quad \text{(F.12)}$$

$$Q_{3,i} = N_{1,2}\left(\Gamma^{t+1} m_{1,i}^{t+1} + m_{2,i}^{t+1}\right) - N_{2,2}\left(\Lambda^{t+1} m_{1,i}^{t+1} + m_{3,i}^{t+1}\right), \quad \text{(F.13)}$$

$$N_{1,1} = 2c\Lambda^t \omega_z^t - c\omega_x^t, \quad \text{(F.14)}$$

$$N_{2,1} = 2c\Gamma^t \omega_z^t, \quad \text{(F.15)}$$

$$N_{1,2} = 2c\Lambda^{t+1} \omega_z^{t+1} - c\omega_x^{t+1}, \quad \text{(F.16)}$$

$$N_{2,2} = 2c\Gamma^{t+1} \omega_z^{t+1}. \quad \text{(F.17)}$$

for $i = 1, 2, 3$ and time t. Once the velocity has been computed, the linear acceleration can be extracted from:

$$\mathbf{V}_{t+1} = a \cdot \mathbf{V}_t \quad \text{(F.18)}$$

and, therefore,

$$a = \frac{N_{2,2}}{V_x}\left(\Gamma^{t+1} m_{1,1}^{t+1} + m_{2,1}^{t+1}\right) + V_y\left(\Gamma^{t+1} m_{1,2}^{t+1} + m_{2,2}^{t+1}\right) + V_z\left(\Gamma^{t+1} m_{1,3}^{t+1} + m_{2,3}^{t+1}\right). \quad \text{(F.19)}$$

Applying the transformation matrix (either $\mathbf{M}(t)$ or $\mathbf{M}(t+1)$), the system can now obtain the velocity vector $\mathbf{v}$ in the eye coordinate space. This allows to solve for all other unknowns, including distance to the point.

References

[1] A. R. Bruss and B. K. P. Horn. Passive navigation. In *Computer Vision, Graphics, and Image Processing*, number 21, pages 3–20, 1983.

[2] H. B. Barlow. Tree theories of cortical function. In R. D. Freeman, editor, *Developmental Neurobiology of Vision.* Plenum Press, New York, 1979.

[3] I. Hadani M. Gur A. Z. Meiri D. H. Fender. Hyperacuity in the detection of absolute and differential displacements of random-dot patterns. In *Vision Research*, number 20, pages 947–951, 1980.

[4] D. Marr and T. Poggio. A computational theory of human stereo vision. In *Proc. of the Royal Society of London B*, number Vol 204, pages 301–328, 1979.

[5] R. Ditchburn. *Eye Movements and Visual Perception.* Clarendon Press, Oxford, 1973.

[6] E. R. Kandel and J. H. Schwartz. *Principles of Neural Science.* Elsevier, New York, 1985.

[7] D. H. Fender. Torsional movements of the eyeball. In *Br. J. Ophthalm.*, number 39, pages 65–72, 1956.

[8] J. J. Gibson. *The Senses considered as Perceptual Systems.* Houghton Mifflin, Boston, 1966.

[9] J. J. Gibson. *The Ecological Approach to Visual Perception.* Houghton Mifflin, Boston, 1979.

[10] H. C. Longuet-Higgins and K. Prazdny. The interpretation of moving retinal images. In *Proc. R. Soc. London B*, number 208, pages 385–387, 1980.

[11] H. Collewijn, R. M. Steinman, C. J. Erkelens, Z. Pizlo, E. Kowler, and J. Van Der Steen. Binocular gaze control under free-head conditions. In H. Shimazu and Y. Shinoda, editors, *Vestibular and brain stem control of eye, head and body movement.* Japan Scientific Society Press, Springer Verlag, 1991.

[12] I. Hadani. The corneal lens goggles and visual space perception. In *Applied Optics*, number 30, 28, pages 4136–4147, 1991.

[13] I. Hadani. Space perception with normal and prosthetic vision. In *SID Digest*, number 23, pages 294–297, 1992.

[14] K. P. H. Horn. *Robot Vision.* The MIT Press, Cambridge, Massachusetts, 1990.

[15] I. P. Howard. Spatial vision within egocentric and exocentric frames of reference. In S. R. Ellis, editor, *Pictorial Communication in Virtual and Real Environments*, pages 338–358. Taylor & Francis, London, 1991.

[16] I. Hadani, A. Kononov, G. Ishai, and H. L. Frish. Two metric solutions to three-dimensional reconstruction for an eye in pure rotations. In *J. Opt. Soc. Am.*, number 11, 5, pages 1564–1574, May 1994.

[17] I. Hadani, A. Z. Meiri, and M. Gur. The effects of exposure duration and luminance on the 3- dot hyperacuity task. In *Vision Research*, number 24, 8, pages 871–874, 1984.

[18] I. Hadani and B. Julesz. Interpupillary distance and stereoscopic depth perception. In *Invest. Ophthalm. and Visual Sci.*, number 33, 4, 1992.

[19] I. Hadani and B. Julesz. Perceptual constancy and the mind's eye looking through a telescope. In *Perception*, 1993. (in progress).

[20] I. Hadani and E. Barta. The hybrid constraint equation for motion extraction. In *Image and Vision Computing*, number 7, 3, pages 217–224, 1989.

[21] I. Hadani, G. Ishai, and B. Julesz. The autokinetic movement and visual stability. In *Invest. Ophthalm. and Visual Sci.*, number 32, 4, page 900, 1991.

[22] I. Hadani, G. Ishai, and M. Gur. Visual stability and space perception in monocular vision: mathematical model. In *J. Opt. Soc. Am.*, number 1, pages 60–65, 1980.

[23] J. J. Koenderink and A. J. van Doorn. Method of stabilizing the retinal image. In *Applied Optics*, number 13, 4, pages 955–961, 1974.

[24] J. J. Koenderink and A. J. van Doorn. Local structure of movement parallax of the plane. In *J. Opt. Soc. Amer.*, number 66, 7, pages 717–723, 1976.

[25] J. J. Koenderink and A. J. van Doorn. Affine structure from motion. In *J. Opt. Soc. Amer.*, number 8, 2, pages 377–385, 1991.

[26] J. Walraven, C. Enroth-Cugell, D. C. Hood, D. I. A. MacLeod, and J. L. Schnapf. The control of visual sensitivity: Receptoral and postreceptoral processes. In L. Spillmann and J. S. Werner, editors, *Visual Perception: The Neurophysiological Foundations*. Academic Press, San Diego, 1990.

[27] J. J. Koenderink. Some theoretical aspects of optic flow. In R. Warren and A. H. Wertheim, editors, *Perception and Control of Self-Motion*, pages 53–68. Erlbaum, Hillsdale, NJ, 1990.

[28] L. L. Sutro and J. B. Lerman. Robot vision. Technical Report Internal Report R-635, Charles Stark Draper Laboratory, Cambridge, Massachusetts, April 1973.

[29] H. C. Longuet-Higgins. A computer algorithm for reconstruction a scene from two projections. In *Nature*, number 293, pages 133–135, 1981.

[30] A. Z. Meiri. On monocular perception of 3-d moving objects. In *IEEE Trans. Pattern Anal. Mach Intell.*, number PAMI-2, pages 582–583, 1980.

[31] H. H. Nagel. On the derivation of 3d rigid point configurations from image sequences. In *IEEE Conference on Pattern Recognition and Image Processing*, 1981.

[32] H. K. Nishihara. Prism: A practical real-time imaging stereo matcher. In *Proc. SPIE Cambridge Symp. on Optical and Electro-Optical Engineering*, number 6-10 November, Cambridge, Massachusetts, 1983.

[33] K. Prazdny. Determining the instantaneous direction of motion from optical flow generated by a curvilinear moving observer. In *Computer Graphics and Image Processing*, number 17, pages 238–248, 1981.

[34] R. Y. Tsai and T. S. Huang. Uniqueness and estimation of 3-d motion parameters and surface structures of rigid objects. In W. Richards and S. Ullman, editors, *Image Understanding 1985-1986*, chapter 6. Albex, Norwood, NJ, 1985.

[35] S. Hecht and E. U. Mintz. The visibility of single lines at various illuminations and retinal basis of visual resolution. In *Journal of General Physiology*, number 22, pages 593–612, 1936.

[36] S. Ullman. *The Interpretation of Visual Motion.* MIT Press, Cambridge, 1979.

[37] G. von Bekesy. *Experiments in Hearing.* McGraw-Hill, New York, 1960.

[38] H. L. F. von Helmholtz. *On the Sensations of Tone.* Dover, New York, 2nd english edition, 1954 edition, 1877.

[39] H. von Schelling. Concept of distance in affine geometry and its application in theories of vision. In *J. Opt. Soc. Am.*, number 46, pages 309–315, 1956.

[40] Y. Trotter, S. Celebrini, B. Stricanne, S. Thorpe, and M. Imbert. Modulation of neural stereoscopic processing in primate area v1 by the viewing distance. In *Science*, number 257, pages 1279–1281, 1992.

[41] L.R. Young. *A sampled data model for eye tracking movements.* PhD thesis, Dept. of Aeronautics and Astronautics, M.I.T, 1962.